ESTIMATING CLASS PEER EFFECTS IN BRAZILIAN PRIMARY SCHOOLS EMPLOYING A REGRESSION-DISCONTINUITY DESIGN

Martin Foureaux Koppensteiner*

This version: May, 2010
Work in progress, please do not cite.

Abstract

I propose a novel way of identifying peer group effects employing a regression-discontinuity design that makes use of the assignment mechanism of students into classes in Brazilian public primary schools. I find that being in the class with older peers deteriorates math test scores at 5^{th} grade significantly by around 0.4 of a standard deviation. Information on the difference in the peer composition suggests that behavioural differences in the two classes adds to simple spillover effects of being with less able peers. Information from a background student questionnaire also suggests that teaching practices may be impeded by greater heterogeneity in age and achievement in the older class.

JEL: I20, I21
Keywords: Peer effects, regression discontinuity, educational production.

* Economics Department, Queen Mary University of London
Correspondence address: Queen Mary, Queens Building, Mile End Road, London E1 4NS, UK.
Email: m.koppensteiner@qmul.ac.uk. Telephone: +44 207 882 8843.

1. INTRODUCTION

The question whether the composition of the peer group matters for the outcome of an individual group member has recently received considerable attention in many contexts where social interactions can be found. Peers have been studied in their effect in the context of schools, universities, in the work place, in local neighbourhoods, and prisons among other on a wide range of outcomes.[1] Education provides an interesting study field for peer effects due to the apparent group character in schools and classrooms and the potential of education policies in changing the peer group composition.[2]

The identification of peer effects is nevertheless difficult due to conceptual problems as well as data limitations. Adhering to the education example, an identification strategy for peer effects needs to address potential endogenous selection of students into the peer groups at the school and class level. With selection into peer groups, unobserved characteristics such as ability, parental support or student effort are likely to be correlated among peers and educational outcomes are therefore correlated within the peer group even in the absence of externalities caused by peer effects.[3] In addition, the analysis needs to deal with separating peer effects from common shocks to the peer group, such as differential pedagogic and teacher inputs and it needs to account for simultaneous determination of student and peer achievement (Manski 1993, Hanushek et al. 2003).

Estimating peer effects using cross-sectional variation in peer characteristics and outcomes is not very promising, as the formation and hence the variation in peer groups are subject to selection into schools and classes.[4] Variation of mean peer characteristics between classes in the same cohort needs to be treated with caution, as class composition may be based on observed and unobserved characteristics of the students. Previous research has approached the selection problem by either controlling for observable characteristics, by estimating selection models, or by comparing siblings of families that move homes and are therefore subject to different schools (Hoxby 2000a). These

[1] Recent studies include Mas and Morretti (2009) for productivity effects on supermarket cashiers, Bandiera, Barankay and Rasul (2010) on social networks and worker productivity in farm production, Guryan, Kroft and Notowidigdo (2009) on the productivity of professional golf players, Bayer, Hjalmarsson and Pozen (2009) on the effect of juvenile offenders serving time on other's subsequent criminal behaviour to name just a few.

[2] Studies on peer effects in education include Hoxby (2000a) for gender and race peer effects, Hanushek et al (2003) provide a framework for estimating peer effects trying to overcome omitted variables and simultaneous equation biases, Duflo, Dupas and Kremer (2008) provide evidence from a randomized experiment in Kenya, Lavy, Paserman and Schlosser (2008) on ability peer effects and potential channels, Lavy, Silva and Weinhardt (2009) on distributional effects of ability peer effects, Lavy and Schlosser (2007) on gender peer effects and their operational channels, Zimmerman (2003) and Sacerdote (2003) for peer effects in college education, Angrist and Lang (2004) for peer effects on racial integration and Ammermueller and Pischke (2006) for a cross-country comparison of peer effects at primary school level. Student tracking, school choice, busing, admission policies, class formation, repetition policies, and residential location decisions are relevant policy issues that can change the peer composition at school and classrooms (Zimmerman 2003 and Hanushek et. al 2003).

[3] See Burke & Sass (2008).

[4] The choice between (public and private) schools in Brazil depends strongly on the socio-economic status of the parents, so that between-school variation is contaminated by selection. Also, access to different quality public schools is often by neighbourhood, which again is self-selected or determined by income and socio-economic status.

methods nevertheless prove rather unconvincing as they do not offer credible ways of dealing with unobservables or they rely on very restrictive identifying assumptions. Randomized experiments seem to be the first choice overcoming the selection problem and there is some recent evidence on peer effects using experimental random assignment of peers by Duflo, Dupas and Kremer (2008) on ability grouping and Whitmore (2005) on gender peer effects in higher education. Empirical strategies using natural experiments, such as conditional random assignment of college roommates by Zimmerman (1999) and Sacerdote (2000), or the use of idiosyncratic gender and race variation on the cohort level by Hoxby (2000a) also have been proposed.[5]

There is still relatively little quasi-experimental evidence that overcomes these important problems in the identification of peer group effects in education. This paper provides quasi-experimental evidence from exogenous variation in the composition of the peer group by using the assignment of students into classes which provides for a natural case of a regression discontinuity (RD) design. In some of the primary schools of the sample students are allocated to classes using their relative age in the cohort as assignment criterion. Using a continuous assignment variable this creates a discontinuity in the assignment to a classroom (peer group). This is a novel approach in the identification of peer effects that helps to overcome the outlined selection problem. Similarly to the prevalent use of RD designs in estimating treatment effects, where treatment is determined by whether the observable forcing variable exceeds a known threshold, the mechanism that assigns students to classes according to their age rank in the cohort, creates treatment variation that can be "as good as random" for individuals close to the class cut-off point.[6] The proposed identification strategy differs nevertheless in some dimensions from typical uses of the RD design. Rather than estimating the effect from a predefined homogenous treatment/programme, treatment varies in the present case from school to school through the assignment to classes with varying mean characteristics.

As the identification strategy is nested within schools and the dataset allows to control for a wide set of class-level characteristics, in particular teacher characteristics, I can quite confidently rule out the possibility that the results are driven by correlated effects.

Employing two-stage-least squares to estimate the discontinuity in a fuzzy RD setup I find strong evidence for peer group effects. I estimate a negative effect of being in the relatively older class in the size of around 0.4 of a standard deviation in math test scores. It is challenging to clearly disentangle the different mechanisms through which the composition of the peer group creates this

[5] An additional difficulty in estimating peer effects can arise from the simultaneity effect, known as reflection problem and outlined by Manski (1993). The reflection problem arises as student i's outcomes may not only be affected by endogenous and exogenous peer effects, but may affect simultaneously outcomes of peers as well. I nevertheless do not regress student i's outcome on peers' outcomes, so that the reflection problem is not relevant in this context.

[6] See Lee and Lemieux (2009) for a comprehensive listing of RD applications in economics.

strong negative effect. I will use information on peer and teacher behaviour to explore some of the possible mechanisms through which peer effects operate in this context.

The remainder of this paper is organized as follows: Section 2 briefly describes the Brazilian educational system and the educational system in Minas Gerais. Section 3 describes the data. Section 4 presents the assignment mechanism of students into classes and introduces the identification strategy. Main results are presented in section 5. Section 6 presents tests for non-random sorting and for correlated effects and section 7 presents an interpretation of the peer group estimates. Section 8 concludes.

2. THE EDUCATIONAL SYSTEM IN BRAZIL AND MINAS GERAIS

Primary schooling is compulsory in Brazil for children from 6 years of age, and consists of nine years of schooling. Children that turn six by 30th June of a given year are required to enrol at primary school.[7] Brazil has a largely decentralized education system. Public schools are either under the administrative control of each state's Secretariat of Education (SEE) or under the control of municipal authorities. The federal administration is left with the responsibility for coordinating educational policies working in an articulated fashion with State and Municipal Education Secretariats and monitoring the comprehensive system of educational funds between the federal and state level. State schools account for more than half (55%) of all public schools and the vast majority of public schools in this analysis are in urban settings (91%). Allocation of students to public schools is based on the area of residence in such a way that parents cannot choose particular schools for their children within the system of public education. There exists a sizeable private sector engagement in the provision of primary schooling, but as private institutions do charge substantial fees, access to private schools is limited to children from middle- and high-income families.[8] Public schools are free of charge at all ages.

In the 2006 wave of the Programme for International Student Assessment (PISA) Brazilian students at age 15 rank in the bottom end of all countries tested with a mean in math test scores of 393 far below the OECD average of 492, which shows that despite rapid improvements during the last decade there are quality concerns with the provision of primary (and secondary) schooling in Brazil (OECD 2006).

[7] The Brazilian school year coincides with the calendar year. See also the data annex for the creation of the student age variable.

[8] Around 10% of school children in Minas Gerais attend private schools. Source: School Census 2007.

3. DATA AND DESCRIPTIVE STATISTICS

As an outcome measure of educational production I use standardized test scores in mathematics of primary school students at 5th grade in public schools in the state of Minas Gerais, the second most populous state in the South-east of Brazil. Educational standards in Minas Gerais are among the highest compared to other states in Brazil.[9] The data used for the analysis comes from two data sets that are linked by school and class identifiers.

In 1999 the SEE has started to build up the State System for the Evaluation of Public Education (SIMAVE), which includes the Programme of Evaluation of Basic Education (PROEB), focusing on the evaluation of student performance in primary and secondary school. The standardized math test score data stems from PROEB and for this study I use the wave of 2007 as it contains the most detailed information on student age on the month level, compared to previous waves of the test. The test is carried out at all public schools in the state of Minas Gerais (which include state and municipal schools) and test scores are standardized to a mean of 500 with a standard deviation of 100. All classes of a given grade at each school are examined and participation is compulsory on the school and individual level leading to a high participation rate of 93% of all students. All pupils answer a detailed socio-economic questionnaire, which includes information on sex, month and year of birth, ethnic background and detailed information on the socio-economic background of the family. Table 1 presents summary statistics of these variables. Average age of students at the test taking date is 11.27 years, which is about ¾ of a year above the appropriate age of this grade. This age-grade mismatch is due to a combination of late enrolment and grade repetition, mostly at third grade.[10]

PROEB also includes a headmaster and teacher questionnaire. The headmaster questionnaire includes questions on personal characteristics of the headmaster, such as age, sex and educational background and questions on school characteristics and pedagogic decisions at the school. The teacher questionnaire includes questions on personal characteristics, as well as teacher evaluations of the class and the students.

The second source of data comes from the 2007 school census in Brazil that is conducted by the National Institute for the Study and Research on Education (INEP) for the Federal Ministry of Education (MEC) and comprises detailed information on school characteristics. It compiles data from all primary schools in Brazil in cooperation with the states' secretariats of education and the municipal authorities. Summary statistics for the schools used in this analysis are presented in table A1 in the annex.

---

[9] In the nation-wide school evaluation system (SAEB) mean performance of pupils from Minas Gerais is clearly above the Brazilian average, ranking 4th highest (INEP 2007).

[10] More details can be found in the data annex.

I focus on schools with two classes per grade for ease of inference on class sorting. The data comprises 16,031 students from 363 public primary schools. Students at these schools are overwhelmingly from deprived socio-economic family backgrounds. The families of 47% of students at these schools receive Bolsa Família, the Brazilian conditional cash transfer programme for poor and very poor families, compared to around 25% of the overall population.[11]

4. EMPIRICAL STRATEGY

4.1 Assignment of students into classrooms

In this paper I combine the mechanism to divide enrolment cohorts based on a maximum class-size cap with the assignment rule of primary school students into classes in Brazilian primary schools to identify class peer effects on individual student's performance. The RD design using the discontinuity induced by a maximum class-size rule has first been proposed by Angrist and Lavy (1999) to estimate the effect of class-size on student proficiency.[12] They use *Maimonides' maximum class-size rule* of 40 students that creates a discontinuous relationship at multiples of 40 in the total number of students in the cohort to estimate class-size effects. Different from Angrist and Lavy I do not use the discontinuous relationship between the cohort size and class-size at the class-size cap using schools with cohort sizes close around multiples of the maximum class-size number. Instead I make use of the discontinuities in the assignment of students to either of the two classes created by the combination of a (school-specific) class-size cap and the age ranking of students to identify peer group effects on students close to the class-caps.

In the setting of public schools of Minas Gerais, when the number of students per entry cohort plus potential repeaters in grade one exceeds multiples of 25 students per class[13], the student cohort is to be divided into the appropriate number of classes. At exact multiples of 25 this theoretically creates a straightforward relationship between the enrolment cohort size and the number of classes with exactly 25 students.[14] At enrolment cohort sizes different from exact multiples of 25, there is some flexibility of the school administration on whether to form equally sized classes. In the case of two classes per school, below an enrolment cohort size of 50, the cohort not necessarily needs to

---

[11] Families are eligible for Bolsa Família, if per capita family income is not above R$ 120 (in 2007) and receive monthly R$ 20 per child under the condition of regular school attendance and participation in vaccination campaigns. Families below a per capita income R$ 60 receive an additional basic family allowance of R$ 62. See http://www.mds.gov.br/bolsafamilia/ and Lindert et al. (2007) for details.

[12] Although the use of the discontinuity to estimate class-size effects is not uncontested for all school systems. See Urquiola and Verhoogen (2009) for a discussion.

[13] Law 16.056 of 24th April 2006 limits class size to 25 students in the initial years of fundamental education (1st-5th grade) in all public schools in Minas Gerais. Exceptions are theoretically only allowed in extenuating circumstances and during the transitory period of the introduction of the law.
http://crv.educacao.mg.gov.br/sistema_crv/banco_objetos_crv/%7B103FA0DB-B47A-4E66-A719-402B21F94D5B%7D_lei%2016056%202006.pdf

[14] I provide more details on the mechanism in the annex.

be divided in equal size classes, leading to a potentially endogenous class-size threshold. In fact I find that older classes are on average 2.58 students smaller than the younger classes (mean size 22.85 and 25.43, respectively).[15]

Furthermore, the class-size cap is not in all cases strictly enforced which also contributes to the flexibility in deciding on real class-sizes away from the predicted class-size of the maximum class-size rule. There is a substantial number of classes exceeding the class-size cap of 25. Around a third of all classes exceed the class-size cap (31%). Only few classes (9%) nevertheless exceed the class size cap by more than 10% and 97.5% of all classes are smaller than 33 students.

With two (or more) classes per grade the school administration needs to decide how to assign students into classes at point of enrolment at first grade. The allocation of students into classes can be done in an arbitrary way by randomly allocating students into classes. In the present case of Brazilian primary schools, in which the age variation within cohorts of students is considerably larger compared to other countries, allocation of students of one cohort at first grade into classes according to age is an important option for the assignment of students.[16] It is often postulated that classes with smaller within-class variation in age make instruction easier for teachers and the education production process more efficient.[17] As age of students at the point of enrolment in first grade can be easily observed by school administrators, differently from innate ability or other behavioural characteristics, age sorting provides a convenient way of grouping students along observable characteristics.

School headmasters in Minas Gerais are free in choosing the allocation mechanism at their school. The assignment mechanism that uses a smooth function in age to naturally order students in a given cohort creates a discontinuity in class membership at the actual class-size cap of the younger class. Figure 1 provides a visual representation of the discontinuity in the mean class rank, where I plot local averages of the class rank according to the individual age rank of students represented as distance from the discontinuity point and local linear regression lines are superimposed.[18]

[15] Lazear (2001) points out that optimal class size varies directly with student behaviour and classes with more disruptive students are often found to be smaller. See Lazear (2001) for a theoretical behavioural model on class-size choice and West and Wößmann (2006) for an empirical analysis on student sorting and endogenous class-size.

[16] There is an extensive pedagogic literature on age, ability grouping, and academic tracking, but little work from economists on theoretical foundations or empirical analyses of the impact of the different forms of forming classes on the mean and the distribution of performance. See Robinson (2008), Adams-Byers, Squiller Whitsell and Moon (2004), Betts and Shkolnik (1999) for some recent examples. Kremer (1997) provides a general economic model of sorting.

[17] See Hoxby and Weingarth (2006) for a discussion. Grouping students according to their age may in fact at least partially coincide with grouping according to ability, as ability likely is correlated with age at time of primary school enrolment. See Cascio & Schanzenbach (2007) and Angrist & Krueger (1991) for a discussion of age effects on educational outcomes.

[18] With the younger class given a value of 0 and the older class a value of 1. The local linear regression lines are fitted separately on both sides of the threshold using a rectangular kernel function with a bandwidth=3. The values are centred controlling for school fixed effects.

As outlined above, the division of an enrolment cohort will not necessarily follow a strict rule that equally divides the students in the cohort over the two classes. The discontinuity point at the class-cap is therefore potentially endogenous, which might cause concerns for using such a discontinuity point for identification in a RD design. I will test for the plausibility of the identifying assumptions later and I will show that non-random sorting around the discontinuity point or the strategic choice of the exact threshold are unlikely to cause a threat to the identification strategy employed. I use actual class size of the younger class to determine the discontinuity point between the two classes and denote the rank of the student at the school specific class-cap of the younger class as $\overline{\mathrm{N}}$.[19]

4.2 Regression discontinuity design

The identification strategy exploits the discontinuity in the assignment rule of students in schools with two classes. As outlined above the treatment assignment mechanism depends on the value of an observed and continuous variable, the age rank n of the individual student in each school in such a way that the probability to receive treatment is a discontinuous function of that variable at the class-size cap $\overline{\mathrm{N}}$. Identification of the treatment effect arises from the fact that just below and above the known cut-off point individuals are very similar in observable and unobservable characteristics, but are part of very different peer groups. The important difference of the regression discontinuity approach from randomization is that while randomization guarantees that treatment and control group are as similar as possible with respect to characteristics other then the treatment variable, the regression-discontinuity design makes treatment and control group very different in the mean values of age and other student characteristics on the class level (van der Klaauw 2002).

As there are assignment imperfections due to the above outlined reasons, the assignment to treatment depends for the individual close enough to the threshold on n in a stochastic manner, but in such a way that the probability of treatment has a discontinuity at $\overline{\mathrm{N}}$ that varies between schools. This leads to the case of a fuzzy regression-discontinuity design, where the size of the discontinuity is smaller than one.[20]

Consider a simple reduced-form model of an education production function

(1) $$Y_{is} = \delta_0 + \delta_1 T_i + f(n) + \varepsilon_i$$

[19] Potentially endogenous discontinuity points have been used for regression discontinuity strategies elsewhere in the literature; see for example Card, Mas & Rothstein (2008).

[20] Furthermore, I pool all schools together in the analysis irrespective of the choice of the allocation procedure chosen.

where Y_{is} denotes the outcome variable math test score for individual i in school s, and T_i is the treatment indicator that takes a value of 0 for individuals in the younger class and 1 for individuals in the older class, ε_i is an individual unobserved error component; ignoring at this stage any covariates one might want to include in the specification to reduce sampling variability in the estimator. Educational achievement measured in test scores, depends on a smooth function $f(\cdot)$ representing the age rank of student i in the cohort that is constructed using the age of individual students, and on being in either the younger or older class indicated by T_i. I employ the regression-discontinuity design to estimate δ_1, the coefficient of interest using the discontinuity at the class cap as an instrument for treatment T_i (being in the older class).
In a first stage-equation I assume that T_i is smooth function of age rank of students in the cohort and a dummy for being above or below the school-specific discontinuity point $\overline{\mathrm{N}}$ given by the maximum class-size rule, D_{is}.

$$(2) \qquad T_i = \gamma_1 + \gamma_2 D_{is} + f(n) + \nu$$

where ν is an error component.
For identification of the class peer effect δ_1, a continuity assumption needs to be satisfied, such that student achievement varies continuously with the forcing variable of the age rank in the cohort, outside of its influence through treatment T_i (Lee & Lemieux 2009), such that assignment to either side of the discontinuity threshold is as good as random. I estimate the above first- and second-stage equations by OLS and the discontinuity by 2SLS, modelling $f(n)$ as a low-order polynomial parametrically.
Public knowledge of the allocation mechanism and the alleged advantage of treatment may invalidate the above continuity assumption crucial to the regression-discontinuity design if, because of that, the forcing variable is subject to manipulation by optimizing agents (McCrary 2008). In the present context there is potential for manipulation of the forcing variable by two sets of agents involved, the parents of the school children and school administrators. If either parents or school administrators are able to manipulate the rank of a student precisely, the "as good as random" assignment to either side of the threshold may fail. Invalidation of the conditions for the consistency of the regression-discontinuity design nevertheless requires precise control over the forcing variable. Starting with the parent's case two forms of manipulation involving the forcing variable may invalidate the above assumption for the RD design. Parents theoretically could manipulate the declared age of their child disclosed to the school administration. To place their

child into a specific class at time of first grade enrolment, parents need to have knowledge of the age distribution of the other students in the entry cohort and of the cut-off point $\overline{N}$. Even if parents were successful in placing their child in their preferred class by manipulating the declared age, this invalidates the assumption for the RD only if parents have precise control over the resulting age rank and place their child exactly at the cut-off point.[21] Of much greater relevance than manipulating the forcing variable in that way, is the potential manipulation by particularly committed parents exerting pressure on school administrators to assign their child to the younger class at initial enrolment or at a later stage. The consequence of a reassignment of a student from the older class into the younger class are much more severe to the identification strategy, as misplaced students automatically rank at the cut-off point in the class-specific age-rank, which would automatically fulfil the *complete manipulation* case necessary for invalidating the RD assumption of continuity.[22]

McCrary (2008) suggests a test for the failure of the random assignment assumption by testing for a discontinuity in the density of the forcing variable around the discontinuity threshold. As the forcing variable in the present case is nevertheless uniformly distributed due to its nature of a relative rank, this test will not be informative in this analysis.[23] If students were strategically re-allocated from the older to the younger class, average age at the cut-off point would reveal a peak; likewise if students from the younger class were selected out to the older class, approaching the threshold from the right, mean age of students would slope down. If the selection of students were then related to performance, this would impair the validity of the RD design. I can test for any such jump in mean age of students around the threshold to test for strategic reassignment of students around the discontinuity point. Furthermore, I test for the balancing properties of a wide range of pre-determined individual variables. A measured discontinuity in the distribution of baseline individual and family characteristics may also be an indication for manipulation, as these observable characteristics are likely related to the effort of parents for manipulation (van der Klaauw 2008).

From the point of view of school administrators, another issue is important for the validity of the proposed identification strategy. The discontinuity threshold $\overline{N}$ is (partially) under the control of

[21] There is nevertheless no advantage to the students of being placed exactly at the cut-off point; on the contrary, for parents to be sure of placing their child in the younger class, strong underreporting of the real age is likely. Furthermore, the enrolment process at first grade involves some form of official identification, so that manipulation of the age of the child is further impeded. Also, if several parents attempt to manipulate the age rank of their child, this most likely affects the predetermined cut-off point $\overline{N}$ as a function of n, inevitably resulting in imprecise control over the forcing variable.

[22] This is true for re-allocating students from the older class into the younger class and vice versa. Being then the oldest or youngest in the class rank, these students will automatically rank on either side next to the threshold.

[23] This holds true depending on cohort size and the distribution of age for age ranks relatively close to the threshold.

the school administration such that it may be shifted along the age rank of students to allocate students into either of the two classes without breaking the age ranking of the students. If for a example given a pre-selected threshold the school administration would like to include the youngest student of the older class rather into the younger class based on some observable characteristics, the cut-off point could simply be shifted by one more rank upwards. In reality this is unlikely to happen as the allocation of students has to be decided before initial enrolment at first grade, so that the school administration likely has no information, say on the performance, of the student other than administrative information such as age.[24] If the selection of the cut-off point by the school administration took into account any observable characteristics of students, this would lead to a jump in any pre-determined characteristics at the cut-off point. The examination of baseline covariates will therefore be an important exercise to test the RD validity.

5. MAIN ESTIMATION RESULTS

In Figure 2 I plot local averages of math test scores and the local linear regression lines on both sides of the cut-off point at which the clear discontinuity in math test scores is apparent. Table 2 presents the first-stage estimates for the size of the discontinuity in the mean class rank, the OLS estimates for the size of the discontinuity in test scores at the discontinuity point $\overline{N}$ and instrumental variable estimates for the causal effect of being moved from the younger class to the older class just around the discontinuity threshold. All specifications include school-fixed effects that account for observed and unobserved differences between schools.[25] Standard errors are heteroskedasticity robust and adjusted for clustering at the school level. The first column presents the estimates for the models including only a quadratic polynomial in age. Column (2) includes a control for the number of grades repeated by students. This may be important as the reallocation of repeaters might lead to a higher proportion of repeaters in the older class, even close to the threshold.[26] The third column additionally includes the whole set of predetermined individual and family characteristics as covariates. The estimates of column (4) include predetermined teacher characteristics in addition to the other covariates.

[24] Which in fact is used for the class allocation of students. As I outline in detail in the annex, there is some evidence for the endogeneity of $\overline{N}$, as class-size, which is under the control of the school administration of the older class is about 3 students smaller compared to the younger class Although this may systematically affect the learning environment of students in both classes in the form of a correlated effect, this nevertheless does not necessarily violate the above continuity condition required for the consistency of estimator. I will discuss the potential bias of this correlated effect of class size on the estimates later.

[25] Using school fixed effects increases the standard errors of the estimates considerably, as more parameters need to be estimated (less degrees of freedom).

[26] This may happen if repeating students move from the younger class to be the youngest student in the older class, so that the propensity for grade repetition might be slightly increased even for students close to the cut-off point. I will later show that there is nevertheless no statistically significant difference in the propensity to repeat at the threshold.

The top panel of table 2 presents estimates for the first stage regression, where the dependent variable is 1 for students being in the older class and zero otherwise. The estimates for the size of the discontinuity range between 0.451 and 0.467, confirming the jump in figure 1. As expected, the inclusion of any controls does not change the first-stage of the regression-discontinuity estimates very much.

The middle panel of table 2 reports the reduced form estimates from an OLS regression with math test scores as the dependent variable on a dummy equal 1 for being to the right of the threshold. Column (1) reports the raw estimate of the jump in math test scores visible in figure 2. The inclusion of a control for grade repetition only slightly reduces the estimates by around 2%.

The bottom panel of table 2 reports the two-stage-least squares estimates for the class peer effects using the same specifications as for the OLS estimations. The size of the class peer effect, without controlling for repeaters or any other covariates, is around 0.57 of a standard deviation in math test scores and very precisely estimated which points to sizable peer group effects. The inclusion of the repetition control reduces the estimates only very slightly to 0.55 of a standard deviation.

Under the weak identifying assumption for the regression-discontinuity design outlined in the previous section, the results can be interpreted as the causal effect on individuals whose treatment status changes, i.e. who switch from the younger class to the older class as the value of n changes from just to the left of $\overline{N}$ just to the just the right of $\overline{N}$. Students that are close enough to the right of the cut-off are very similar in their characteristics compared to the students just to the left of the cut-off, but are faced with a different peer group composed of a larger share of repeaters, a higher proportion of male classmates, a peer group with a lower overall socio-economic family background and greater heterogeneity in age.

Table 6 presents the RD estimates for wider intervals around the cut-off point and different order of the polynomial included in the regressions. Rows (1) and (2) are the estimates of the RD without any further controls, rows (3) and (4) are the estimates including the full set of controls (including all individual, family and teacher covariates). The estimates do not reveal any substantial sensitivity with respect to the order of the polynomial included. Including a cubic term leaves the estimates virtually unchanged. Increasing the range of observations used for the estimation also does not alter the estimates for the treatment effect in a relevant way.

6. TESTS FOR NON-RANDOM SORTING AND CORRELATED EFFECTS

The key identifying assumption for the estimation in the regression discontinuity design is that around the class cut-off point, all predetermined characteristics of students are balanced on both sides of the discontinuity. Although some discontinuities in average values of pre-determined covariates do not necessarily invalidate the identification assumptions of the regression-discontinuity design, they may at least cast some doubt on the estimation strategy or may be an indication for misspecification of the functional form. I use the rich information from the student questionnaire on individual and family characteristics to formally test the balancing properties of these pre-determined characteristics.

Furthermore, I want to exclude the possibility that the estimated discontinuity is driven by *correlated effects* or common shocks, for example if the learning environment is systematically different for the two classes. This includes the strategic allocation of teachers with different qualities to the different classes, the differential provision of teaching material or different class-size. I therefore test the balancing property also for a range of teacher and some other classroom characteristics.

6.1 Student and family characteristics

Figures 3-22 (letter a) provide an informative graphical analysis of the balancing properties of baseline covariates by plotting local averages for the covariates and fitting local linear regression lines separately on both sides of the threshold employing the same bandwidth and kernel function used for the graphs of assignment and outcome variable for a bin size of one months. All variables are centred to control for school fixed effects, such that the mean deviation from 0 is reported. Figures 3a to 8a show the sex and racial composition of students around the cut-off point. Figure 3a shows that the local proportion of girls reduces smoothly with the age rank leading to a lower mean share of girls in the older class, which can be seen in graph 3b. Letter *b* graphs are the equivalent graphs to letter *a* graphs on the peer values of all variables using a leave-one-out mean in place of the own value. While the proportion of mixed and black students suggest a small jump at the cut-off point, the variables for the proportion of white, East-Asian and indigenous students appear as continuous function in the age rank at the discontinuity point. Figure 9a presents local averages for individual age of students in months. As outlined before, reallocation of students from one class to the other would lead to mean age to peak to the left and dip to the right of the threshold. Mean age is nevertheless a smooth increasing function of the rank as distance from the cut-off point, so there is no evidence for strategic reallocation of students between the two classes. Figures 10a to 22a present the same graphical representation of the local linear regression fits and local averages for a wide range of predetermined parental characteristics. Apart from the number of cars and DVD players per household these variables appear well balanced on both sides of the cut-off point and there is little sign of any considerable discontinuity of these characteristics at the cut-off point. Only the two above mentioned variables suggest a small jump around the discontinuity point. From two additional proxies for the socio-economic status of the family, the number of domestic workers employed and the fraction of families receiving Bolsa Família, only the latter suggests some difference around the threshold.

In a formal analysis I estimate the RD in the same specification as for math test scores above for these variables. Table 3 reports the estimates for these variables at the threshold. Only the RD estimate for the probability of being black is significant on the 5% level, none of the other estimates reveal any statistically significant difference in student characteristics around the discontinuity point.[27] None of the other socioeconomic characteristics of the student households reveals a statistically significant difference at the threshold and most estimates are small, confirming that the balancing properties of these predetermined characteristics are satisfied. The estimate for age reveals the smooth transition across the threshold, and there is no sign of a negative discontinuity suggestive of non-random sorting of students around the threshold.

From the RD estimates of the pre-determined covariates and the inspection of the graphs there is no indication for a discontinuity and non-random sorting of students around the threshold that would impair the identifying assumptions of the regression-discontinuity design. Although the absence of discontinuities in predetermined individual and family characteristics does not prove the balancing property of unobservables, it is reassuring to find that individuals on both sides of the cut-off are observationally identical.[28]

The inclusion of these additional individual and family controls in column (3) of table 2 changes the estimates for the reduced-form regressions nevertheless somewhat. Likewise, the IV estimates of the class peer effect are around 20% smaller than without the inclusion of these controls, leaving some role for individual level heterogeneity in the estimation of the peer group effect. By considering observations that are in a range of one month from the class discontinuity point, this leads to a range of about two months that naturally leads to some small differences in individuals. Shrinking the neighbourhood to reduce the bias is nevertheless not possible, as age is reported on the month level only.[29] As I do not find any discontinuity in the predetermined characteristics at the cut-off point, the reduction in the estimated peer effect may be due to model misspecification when including the large range of controls (Imbens & Lemieux 2008).

[27] Choosing different specifications for the RD by including either only a linear polynomial term or a cubic term makes the estimate for this variable seizing being significant, so that the single significant estimate can either be attributed to model misspecification or random chance. Any other specification for the functional form or estimating the RD without robust standard errors does not change insignificance of the estimates of any of the variables.

[28] Only one of 21 coefficients shows a significant discontinuity, which given the 5% level of significance is well in line with random chance.

[29] In fact the average age differences for observations within one month rank from the cut-off point is slightly above 2 months (2.667 months) and significantly estimated. This is due to some students being further away from the cut-off than one month, but being ranked the closest from the threshold.

6.2 Teacher characteristics and class environment

Another concern for the estimation of class peer effects is that correlated effects in form of common shocks to the peer group can bias the peer effect estimates as discussed above. Common shocks may lead to a bias in the estimation of the peer group effect, if the learning environment for the students in the two classes is systematically different. Although it is difficult to completely rule out the existence of any differences in the learning environments, I can nonetheless assess whether there exist differences in the learning environment of for example teacher characteristics.

As outlined above, such systematically different learning environments may be created by systematically assigning teachers with specific qualities to either of the two classes. This might happen in a compensatory fashion, such that better teachers are allocated to weaker classes (in this case to the older classes with a higher proportion of repeaters), which would lead to an underestimation of the age peer effect. Better educated or more experienced teachers could also be allocated to the younger class to strengthen good students further, which would lead to overestimating the peer effect. This might be the case if the more committed (or more able) parent's of students in the younger more successfully lobby school administrators to receive the more advantageous learning environment than parents from already relatively disadvantaged children in the older class.[30]

Generally, one might like to find pre-determined teacher characteristics to be equal on average in the two classes. I estimate OLS regressions of teacher and class characteristics on a dummy equal one for the older class and table 4 reports the coefficients. For each variable a separate regression has been estimated. None of the teacher characteristics show any significant difference between the two classes, including teacher sex, age, race, experience, teacher education and training. This is very reassuring, as strategic teacher allocation does not seem to play any role. There is no indication that more experienced or better educated teachers are assigned to the younger class, so that the results could at least partially be driven by teacher quality. Neither are those teachers systematically allocated to the older classes with the greater proportion of repeaters, less female students and with students with lower socio-economic status in a compensatory fashion. Including all teacher characteristics in the RD estimates (table 2, column (4)) also does not change the treatment effect for the peer group in any relevant way. Teacher statements about the allocation of teaching resources to either of the classes also provide some more evidence that the results are not biased by common shocks to the class room creating different teaching environments. I use answers from the teacher questionnaire about the quantity of pedagogic resources available for teaching reported on the class-level by the teachers to investigate class level teaching resources.

[30] As nevertheless the allocation of class teachers to the classes is done in first grade and the class teacher stays with the class for most often the first five years of primary school, there is little scope for such action.

None of the variables reported by teachers (frequency of class council meetings, quality of teaching books, the occurrence of insufficient financial resources and pedagogic resources for class teaching) are significantly different between the two groups.

As I have already pointed out, there is some concern about the choice of class-size. The estimate for class-size in table 4 reveals that class-size of the older class is in average about 3 students smaller. As class-size may affect student achievement, this potentially leads to a bias in the estimation of the peer group effect. Although the literature is somewhat inconclusive about the empirical relationship of different class sizes on student achievement, there is some agreement that smaller classes are to be preferred (see Angrist and Lavy 1999 and Urquiola 2006); the effects reported in the literature nevertheless appear to be rather moderate. As in present case the older class is in average smaller, this may lead to a downward bias of the true peer group effect on student outcomes and I may need to take the estimated results as a lower bound for the true effect. Given the smaller class-size of the older class, there is nevertheless no threat that the estimated class peer effect may in fact (at least partially) be due to a class-size effect.

Table 4 also reveals that the proportion of students that do not participate in the PROEB test, either because they are ill on the test date or have any other reason not to be present at the test, differs between the two classes. The non-participation rate is about 7% higher in the older class. This potentially can bias the estimates for the class peer effect; the direction of the bias depends on latent proficiency of the students failing to participate in PROEB. Under the assumption that it is more likely for worse performing students to be absent from the test, this potentially leads to underestimating the true class peer effect. The size of a bias associated with the difference in the participation rate is nonetheless likely negligible. Unfortunately I do not have information on the characteristics of the absent students or the reason for the absence, nor can I identify whether or not an absent student was ranked around the threshold.

7. INTERPRETATION OF THE ESTIMATED PEER GROUP EFFECT

Column (2) of table 3 reports the estimates for the mean peer group variables for the same individuals around the cut-off (the mean values for all other students in the class less the students around the discontinuity point). Some of these characteristics are correlated with the age rank of students (through age), so that their means or proportions increase or decrease (smoothly) across the discontinuity point. It is important to notice as with an increasing bandwidth the differences across the discontinuity point may become significant. This is visualized in figures 3b-23b. Some of the peer variables reveal a clear jump at the discontinuity point. Most clearly this is visible for mean age (figure 9b). But also the proportion of girls in the peer group for students at the discontinuity point (figure 3b), and other mean characteristics reveal a clear albeit modest discontinuity at the cut-off point, pointing out that students on either side of the threshold face very different peer groups. Many of the RD estimates for these variables in table 3, column (2) reveal a significant discontinuity, confirming mean differences in the peer groups in terms of a wide range of socio-economic characteristics including racial background and the sex composition.

All characteristics indicate a lower socio-economic status of the family in the older class. On average, students just to the right of the threshold, although not being different from students just to the left on the whole range of individual and parental characteristics, are in a class with peers from clearly more disadvantaged families. Their peers are on average ¾ of a year older, the proportion of girls in the class is smaller, and the proportion of children with families that receive a cash-transfer through Bolsa Família is higher. The negative estimates on the peer group effect, could partially be explained with simple spillover effects of the more disadvantaged students in the peer group of the older classes, in the interpretation as exogenous effects in the language of Manski (1993). Despite the large difference in average age in the two classes, the differences in the socio-economic composition appear nevertheless rather moderate. The most remarkable difference in the average characteristics of the two groups can be found the age difference. The older class is on average almost a year older than the younger class, which is to a good proportion driven by the larger proportion of repeaters. The estimated discontinuity in the average number of years repeated is 0.6 years larger for students just to the right of the cut-off point. While 79.7% of students in the younger class report to have never repeated, this number drops to only 48.9% in the older class.

Not only does mean age differ greatly between the two classes, but the age distribution between the two classes is also quite different.[31] The standard deviation in age (table 4, row 16) is about 3.5 months larger in the older class. Strikingly, figure 2 reveals furthermore that the achievement distribution between the two classes also differs greatly. While the local linear regression fit is almost flat in the younger classes, the fit in the older class shows a negative correlation between the age rank and test score results. A more heterogeneous class composition along age or achievement has been associated in the literature with impairing teaching practices related to a focus model of peer effects (Hoxby and Weingarth 2006, Lavy, Paserman and Schlosser 2008).

Table 5 presents marginal effects from a linear probability model on student reported teaching practices. In the background questionnaire of PROEB students were asked to report on a range of teaching related items. I have estimated separate regressions for each of the dependent variables reported. The estimates reveal that from a student perspective teaching practices are perceived quite differently between the two classes. As I have shown that there are no observable differences

[31] Graphs A1 and A2 in the annex show the histograms for age of students in months for the two classes.

in teacher quality between the two classes,[32] the findings suggest that teaching practices are affected directly by the different peer group composition. In the older class students report that teacher availability to clarify doubts of students is about 12% lower, teachers less likely explain the course content ensuring that all students have understood the taught material and teachers give less opportunity to students to express themselves in class. From an individual perspective, students in the older class report much more frequently that the teacher helps some students in the class more than others, which is suggestive support for the hypothesis that larger heterogeneity in the class room affects teaching practices. There is also a statistically significant difference in the teacher reported percentage of the scheduled curriculum taught in the older class. Teachers in the older class manage to teach almost 4% less of the scheduled curriculum at 5th grade (table 4, row 24).

Apart from effects of heterogeneity in the classroom on teaching practices, larger variation in age may also contribute to behavioural changes of students in the classroom. Teachers of the older classes report also slightly increased level of disciplinary problems with students, the estimate in table 4 is nevertheless only marginally significant. Mean levels of student reported noise and disruption of teaching confirms nevertheless behavioural differences in the two classes (bottom rows of table 5). The probability for classmates being noisy and disruptive is 12% higher in the older class, teachers need to wait much more frequently at the beginning of the class until they can start teaching, and students leave the classroom 20% more often in the older class. The magnitude of these differences in the two classes suggests that behavioural changes induced by the composition of peer groups may play a significant role in explaining the estimated negative peer group effect for students close to the threshold. It is nevertheless not clear from the analysis whether these behavioural changes are induced by the greater heterogeneity in age and achievement between the two classes or by the higher propensity for individual students in the older class that disrupt class teaching practices in line with a Bad Apple model of peer effects (Hoxby and Weingarth 2006, Lazear 2001).

Note also that the identification strategy used to identify peer effects, will give an estimate of the effect of peer age (and its distribution) for students at the cut-off point, where being with the on average older class also means that the students is the youngest in the class, and being with the on average younger class also means that the student is at the top of the class age distribution.

7. CONCLUSIONS

[32] In support to the previously reported results that teacher characteristics are not systematically different between the two classes, students do not report a statistically significant difference in the frequency with which teachers correct their homework, encourage discipline ("teacher makes students pay attention") and ensure progress of students.

In this paper I introduce a novel way of identifying class peer effects using a continuous rule of assigning students into classes that creates a discontinuity in the assignment to the peer group for students just below and above the class-size cap. I report precisely estimated strong negative average treatment effects of the peer group on standardized math test scores for 5th graders in Brazilian primary schools for students close to the discontinuity point. Switching treatment status from the younger to the older class lowers math test scores at 5th grades for these students by around 0.4 of a standard deviation, as a lower bound of the peer group effect. I interpret this negative estimate as the effect of a spillover in the production of education that may be due to a combination of exogenous differences in the mean peer characteristics across the two classes and behavioural changes in the classroom. There is also evidence for changes in the efficiency of educational production suggested by student differences in teaching practices in the two classes. The estimates may need to be considered as lower bounds of the true effect of the class peer, as smaller class-size and the smaller proportion of test takers in the older class may lead - under some assumptions - to underestimation of the true effect.

I test the balancing properties of pre-determined student and parental characteristics by examining these for discontinuities close to the cut-off point and do not find any systematic discontinuity that may arguably infringe the underlying identifying assumptions. This is particularly reassuring given the potentially endogenous class size rule, as suggested by the difference in class size between the classes. These tests are very reassuring in showing no evidence for manipulation of the assignment rule of pupils into the two classes. Likewise, there is no evidence for strategic behaviour of school administrators in sorting students around the threshold.

In addition, this paper provides an understanding of school responses to the given differences in the quality of students in classes. There is no evidence that the results are driven by strategic choices of the school by systematically assigning teachers with different quality or different characteristics to the two classes or by allocating a different set of resources or providing a different learning environment to the two classes, apart from differences in the class-size of the two classes.

REFERENCES

Adams-Byers, Jan, Sara Squiller Whitsell, and Sidney M. Moon. 2004. “Academic and Social/Emotional Effects of Homogeneous and Heterogeneous Grouping.” *Gifted Child Quarterly*, 48(7): 7-20.

Ammermueller, Andreas, and Jörn-Steffen Pischke. 2006. “Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study.” *Journal of Labor Economics*, 27(3): 315-348.

Angrist, Joshua, and Alan B. Krueger. 1991. “Does Compulsory School Attendance Affect Education and Earnings?” *Quarterly Journal of Economics*, 106(4): 979-1014.

Angrist, Joshua, and Kevin Lang. 2004. “Does School Integration Generate Peer Effects? Evidence from Boston’s Metco Program.” *American Economic Review*, 94(5): 1613-1634.

Angrist, Joshua, and Victor Lavy. 1999. “Using Maimonides' Rule to Estimate the Effect of Class Size on Children's Academic Achievement.” *Quarterly Journal of Economics*, 114(2): 533-575.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul. 2010. **“**Social Incentives in the Workplace**.”** *Review of Economic Studies*, 77(2): 417-459.

Bayer, Patrick, Randi Hjalmarsson, and David Pozen. 2009. “Building Criminal Capital behind Bars: Peer Effects in Juvenile Corrections.” *Quarterly Journal of Economics*, 124(1):105-147.

Betts, Julian R., and Jamie L. Shkolnik. 1999. “The Effects of Ability Grouping on Student Achievement and Resource Allocation in Secondary Schools.” *Economics of Education Review*, 19: 1–15.

Burke, Mary, and Tim R. Sass. 2008. “Classroom Peer Effects and Student Achievement.” *Working Papers 08-5*, Federal Reserve Bank of Boston.

Card, David, Alexandre Mas, and Jesse Rothstein. 2008. “Tipping and the Dynamics of Segregation.” *Quarterly Journal of Economics*, 123(1):177-218.

Cascio, Elizabeth, and Diane Whitmore Schanzenbach. 2007. “First in the Class? Age and the Education Production Function.” *NBER Working Paper* 13663.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2008. “Peer Effects, Class Size, and Teacher Incentives: Evidence from a Randomized Evaluation in Kenya.” *CEPR Discussion Paper No. DP704.*

Guryan, Jonathan, Kory Kroft, and Matthew J. Notowidigdo. 2009. “Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments.” *American Economic Journal: Applied Economics*, 1(4): 34–68.

Hanushek, Eric, John F Kain, Jacob M Markman, and Steven G Rivkin. 2003. “Does Peer Ability Affect Student Achievement?” *Journal of Applied Econometrics*, 18(5): 527-544.

Hoxby, Caroline. 2000a. “Peer Effects in the Classroom: Learning from Gender and Race Variation.” *NBER Working Paper* 7867.

Hoxby, Caroline M., and Gretchen Weingarth. 2006. "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects." *Unpublished Manuscript.*

Kremer, Michael. 1997. "How much does Sorting Increase Inequality," *Quarterly Journal of Economics*, 112(1): 115-139.

Lavy, Victor, Daniele Paserman, and Analia Schlosser. 2008. "Inside the Black Box of Ability Peer Effects: Evidence from Variation in Low Achievers in the Classroom", *NBER Working Paper* 14415.

Lavy, Victor, and Analia Schlosser. 2007. "Mechanisms and Impacts of Gender Peer Effects at School." *NBER Working paper No. W13292.*

Lavy, Victor, Olmo Silva, and Felix Weinhardt. 2009. "The Good, the Bad and the Average: Evidence on the Scale and Nature of Ability Peer Effects in Schools." *Unpublished Manuscript.*

Lazear, Edward P. 2001. "Educational Production." *Quarterly Journal of Economics*, 156(3): 777-803.

Lee, David, and Thomas Lemieux. 2009. "Regression Discontinuity Designs in Economics." *NBER Working Paper* No. 14723.

Lindert, Kathy, Anja Linder, Jason Hobbs, and Bénédict de la Brière. 2007. "The Nuts and Bolts of Brazil's Bolsa Família Program: Implementing Conditional Cash Transfers in a Decentralized Context." *Social Protection Discussion Paper* 0709, World Bank.

Mas, Alexandre, and Enrico Moretti. 2009. "Peers at Work." *American Economic Review*, 99(1): 112-145.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 60: 531-542.

McCrary, Justin 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics*, 142(2): 698-714.

OECD (Organization for Economic Cooperation and Development). 2006. "*Equally prepared for life? How 15-year-old boys and girls perform in school*". OECD. Paris.

Robinson, Joseph P. 2008. "Evidence of a Differential Effect of Ability Grouping on the Reading Achievement Growth of Language-Minority Hispanics." *Educational Evaluation and Policy Analysis,* 30(2): 141–180.

Sacerdote, Bruce. 2003. "Peer Effects with Random Assignment: Results for Dartmouth Roommates." *Quarterly Journal of Economics*, 116: 681-704.

Urquiola, Miguel. 2006. "Identifying Class Size Effects in Developing Countries: Evidence from Rural Bolivia." *Review of Economics and Statistics,* 88(1): 171–77.

Urquiola, Miguel, and Eric Verhoogen. 2009. "Class-Size Caps, Sorting, and the Regression-Discontinuity Design." *American Economic Review*, 99(1): 179–215.

Van der Klaauw, Wilbert. 2002. “Estimating the Effect of Financial Aid Offers on College Enrolment: A Regression-Discontinuity Approach.” *International Economic Review*, 43(4):1249-1287,

West, Martin R., and Ludger Wößmann. 2006. “Which School Systems Sort Weaker Students into Smaller Classes? International Evidence.” *European Journal of Political Economy*, 22(4):944-968.

Whitmore, Diana. 2005. “Resource and Peer Impacts on Girls’ Academic Achievement: Evidence from a Randomized Experiment.” *American Economic Review*, 95(2): 199–203.

Zimmerman, David. J. 2003. “Peer Effects in Academic Outcomes: Evidence from a Natural Experiment.” *Review of Economics and Statistics*, 85(1): 9-23.

TABLE 1
MEANS AND PROPORTIONS OF STUDENT AND TEACHER CHARACTERISTICS

Student characteristics		class 1	s.e.	class 2	s.e.	Difference	s.e.
Sex	Female	0.524	(0.005)	0.458	(0.006)	0.066	(0.008)
Age	(in years)	10.930	(0.009)	11.670	(0.014)	-0.740	(0.016)
Race	White	0.306	(0.005)	0.264	(0.005)	0.042	(0.007)
	Mixed	0.526	(0.005)	0.517	(0.006)	0.009	(0.008)
	Black	0.097	(0.003)	0.143	(0.004)	0.046	(0.005)
	East-Asian	0.027	(0.002)	0.034	(0.002)	-0.007	(0.003)
	Indigenous	0.044	(0.002)	0.042	(0.002)	0.002	(0.003)
Repeater	Never repeated	0.797	(0.004)	0.489	(0.006)	0.308	(0.007)
	Once repeated	0.142	(0.004)	0.292	(0.005)	-0.150	(0.007)
	Twice repeated	0.043	(0.002)	0.148	(0.004)	-0.105	(0.005)
	Repeated 3 or more times	0.018	(0.001)	0.070	(0.003)	-0.052	(0.003)
SES	Families with Bolsa Família	0.480	(0.005)	0.592	(0.006)	-0.112	(0.008)
	HH with domestic aid	0.137	(0.004)	0.113	(0.004)	0.024	(0.005)
Means	Num. of books	23.496	(0.322)	19.428	(0.330)	4.068	(0.463)
	Num. of cars	0.608	(0.009)	0.503	(0.009)	0.105	(0.013)
	Num. of computers	0.262	(0.005)	0.195	(0.005)	0.067	(0.007)
	Num. of fridges	0.999	(0.005)	0.958	(0.006)	0.041	(0.008)
	Num. of freezers	0.302	(0.006)	0.282	(0.007)	0.020	(0.009)
	Num. of radios	1.342	(0.008)	1.286	(0.009)	0.056	(0.012)
	Num. of TV	1.497	(0.008)	1.396	(0.009)	0.101	(0.012)
	Num. of DVD players	0.849	(0.007)	0.786	(0.008)	0.063	(0.011)
	Num. of bathrooms	1.246	(0.006)	1.175	(0.006)	0.071	(0.009)
	Num. of washing machines	0.758	(0.007)	0.752	(0.007)	0.006	(0.010)
	Num. of tumble dryers	0.168	(0.005)	0.163	(0.005)	0.005	(0.007)
Teacher characteristics		**class 1**	**s.e.**	**class 2**	**s.e.**	**difference**	**s.e.**
Sex	Female	0.983	(0.011)	0.965	(0.015)	0.018	(0.013)
Age	(in years)	40.495	(0.468)	40.094	(0.486)	0.401	(0.674)
Race	White	0.456	(0.030)	0.477	(0.030)	-0.021	(0.042)
	Mixed	0.420	(0.029)	0.399	(0.029)	0.021	(0.042)
	Black	0.093	(0.017)	0.081	(0.016)	0.012	(0.024)
	East-Asian	0.028	(0.010)	0.039	(0.012)	-0.011	(0.015))
	Indigenous	0.004	(0.004)	0.004	(0.004)	0.000	(0.005)
Highest	Secondary education	0.100	(0.018)	0.118	(0.019)	-0.018	(0.026)
edu. level	Higher edu – ped. degree	0.210	(0.024)	0.208	(0.024)	0.002	(0.034)
	Higher edu - regular	0.410	(0.029)	0.389	(0.029)	0.021	(0.041)
	Higher edu - licentiatura	0.203	(0.024)	0.174	(0.022)	0.029	(0.033)
	Higher edu – other	0.076	(0.016)	0.111	(0.019)	-0.035	(0.024)
	Salary (in R$)	771.74	(22.803)	743.60	(23.754)	28.14	(32.920)
	Years exp. in education	14.023	(0.360)	13.862	(0.375)	0.161	(0.520)
	Years exp. at this school	8.227	(0.397)	7.257	(0.376)	0.970	(0.547)
	Years exp. with this grade	4.697	(0.152)	4.764	(0.152)	-0.067	(0.213)
	Part. in continued training	0.375	(0.028)	0.363	(0.029)	0.012	(0.040)

Notes: The date from the upper panel is taken from the student background questionnaires, the data from the lower panel from the teacher questionnaires of PROEB.

TABLE 2
MAIN ESTIMATION RESULTS

	(1)	(2)	(3)	(4)
	First stage Dependent variable: class rank			
	0.467***	0.464***	0.453***	0.451***
	(0.056)	(0.055)	(0.057)	(0.056)
R^2	0.326	0.334	0.370	0.403
	Reduced form Dependent variable: math test scores			
	-26.445***	-25.708***	-19.196**	-19.513**
	(7.458)	(7.562)	(7.646)	(7.743)
R^2	0.405	0.420	0.482	0.485
	IV regression discontinuity results Dependent variable: Math test scores			
	-56.574***	-55.461***	-42.385***	-43.297***
	(15.299)	(15.720)	(15.455)	(15.673)
R^2	0.410	0.422	0.485	0.489
Observations:	1,688	1,688	1,688	1,688
Repetition controls	no	yes	yes	yes
Individual controls	no	no	yes	yes
Teacher controls	no	no	no	yes

Notes: The top panel reports the first stage regressions using OLS estimating equation (2). The middle panel reports the coefficient on math test score on the dummy equal 1 for the age rank larger then 0 (reduced form). Test scores are centred using school fixed effects in all specifications. The bottom panel reports IV estimates of the effect of being in the second (the in average older) class on math test scores, where being in the second class has been instrumented by a dummy for having an age rank larger 0. All specifications include a second-order polynomial in age. Specifications in column (2) include a control for the number of times students have repeated a grade, specifications in column (3) additionally include the whole set of predetermined individual and family characteristics, including sex, race and SES family characteristics, specifications in column (4) additionally include all predetermined teacher characteristics, including sex, race, age, salary, variables on educational background and experience. Heteroskedasticity robust standard errors are clustered by schools and reported in parenthesis. **, *** denote significance at the 5% and 1% level, respectively.

TABLE 3

RD ESTIMATES OF PREDETERMINED INDIVIDUAL AND FAMILY VARIABLES

		(1)		(2)	
		Individuals		Peers	
Proportion	Female	0.190	(0.127)	-0.088***	(0.019)
	White	0.008	(0.092)	-0.035	(0.023)
	Black	0.115**	(0.055)	0.089***	(0.018)
	Mixed	-0.037	(0.102)	-0.072**	(0.032)
	East-Asian	-0.026	(0.022)	0.011	(0.009)
	Indigenous	-0.076	(0.047)	-0.001	(0.009)
	Domestic helper	-0.020	(0.058)	-0.053***	(0.017)
	Bolsa Família	0.165*	(0.099)	0.144***	(0.027)
	Age (in months)	0.442	(0.735)	8.157***	(0.796)
	Years repeated	0.061	(0.094)	0.624***	(0.059)
Number of	Bathrooms	-0.101	(0.098)	-0.129***	(0.033)
	Books	-4.314	(4.956)	-8.016***	(1.928)
	Cars	-0.167	(0.138)	-0.141***	(0.039)
	Computers	-0.031	(0.068)	-0.108***	(0.022)
	Fridges	0.096	(0.077)	-0.074**	(0.031)
	Freezers	-0.013	(0.087)	-0.052**	(0.025)
	Radios	0.195	(0.158)	-0.083	(0.052)
	Washing machines	0.080	(0.105)	-0.037	(0.033)
	Dryers	-0.057	(0.082)	0.014	(0.021)
	DVDs	0.125	(0.121)	-0.120***	(0.035)
	TV sets	-0.008	(0.141)	-0.194***	(0.042)
	Video players	0.080	(0.107)	-0.066**	(0.028)
	Number of observations	1,688		1,688	

Notes: Entries are separate IV estimates of the class effect on student and family characteristics, where being in the second class has been instrumented by a dummy for having an age rank larger 0. For each variable a separate regression has been estimated. Column (1) reports the effect around the discontinuity point for the individual values of the characteristics, column (2) reports the estimates for the values of the peer group characteristics for the same individuals around the cut-off point. All regressions include a linear age control and a second-degree polynomial in age. Heteroskedasticity robust standard errors, clustered on the school level are reported in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level, respectively.

TABLE 4
TEACHER AND CLASS CHARACTERISTICS

Dependent variables			
Teacher characteristics	Female	-0.018	(0.017)
	Age (in years)	-0.601	(0.886)
	White	-0.012	(0.060)
	Mixed	-0.000	(0.060)
	Black	-0.004	(0.036)
	East-Asian	0.012	(0.023)
	Indigenous	0.004	(0.006)
	Higher education	-0.018	(0.038)
	Postgraduate degree	-0.016	(0.059)
	Years passed since graduating	0.257	(0.715)
	Teacher salary (in Reais)	-47.321	(38.780)
	Participation in cont. training	-0.015	(0.048)
	Experience in education (in years)	-0.480	(0.686)
	Experience at current school (in years)	-1.048	(0.714)
	Experience with current grade (in years)	0.030	(0.314)
Class characteristics	Std. deviation of age (in months)	3.423***	(0.261)
	Class size	-2.880***	(0.359)
	Nonparticipation rate	0.071***	(0.020)
Class teacher statements	Frequency of class council meetings	0.102	(0.086)
	Quality of books	0.069	(0.054)
	Insufficient financial resources	0.005	(0.057)
	Insufficient pedagogic resources	0.049	(0.070)
	Disciplinary problems with students	0.139*	(0.078)
	% of planned curriculum taught	-3.775***	(0.909)
	% of students to finish primary school	-0.284***	(0.090)
	% of students to finish secondary school	-0.298**	(0.123)
	Number of observations	726	

Notes: Entries are estimates of a linear probability model on a dummy of classrank, where classrank=0 for the younger class and classrank=1 for the older class. For each variable a separate regression has been estimated. The data comes from the teacher questionnaire of PROEB and the school census (for class characteristics). Class teacher statements come from the teacher questionnaire and relate to the specific class taught. Class size is computed using the official number of students enrolled in a class based on information from the school census. Nonparticipation rate is based on the difference between class size and number of students participating in the PROEB test. Heteroskedasticity robust standard errors are reported in parentheses. *, **, *** denote significance at the 10%, 5% and 1% level, respectively.

TABLE 5
TEACHER AND STUDENT CLASS BEHAVIOUR (REPORTED ON STUDENT LEVEL)

Teacher absenteeism	0.122***	(0.036)
Teacher makes students pay attention	-0.029	(0.021)
Teacher corrects homework	-0.031	(0.020)
Teacher ensures students progress	-0.035	(0.022)
Teacher interested in learning progress	-0.084***	(0.022)
Teacher explains until all students understand	-0.082***	(0.027)
Teacher availability to clarify doubts	-0.115***	(0.025)
Teacher gives opportunity to express oneself	-0.086***	(0.031)
Teacher needs to wait to start teaching	0.160***	(0.050)
Teacher helps some students more	0.251***	(0.043)
Student learn taught material	-0.132***	(0.029)
Students are noisy and disruptive	0.115***	(0.044)
Students pay attention in class	-0.027	(0.032)
Students leave classroom early	0.210***	(0.045)
Number of observations	726	

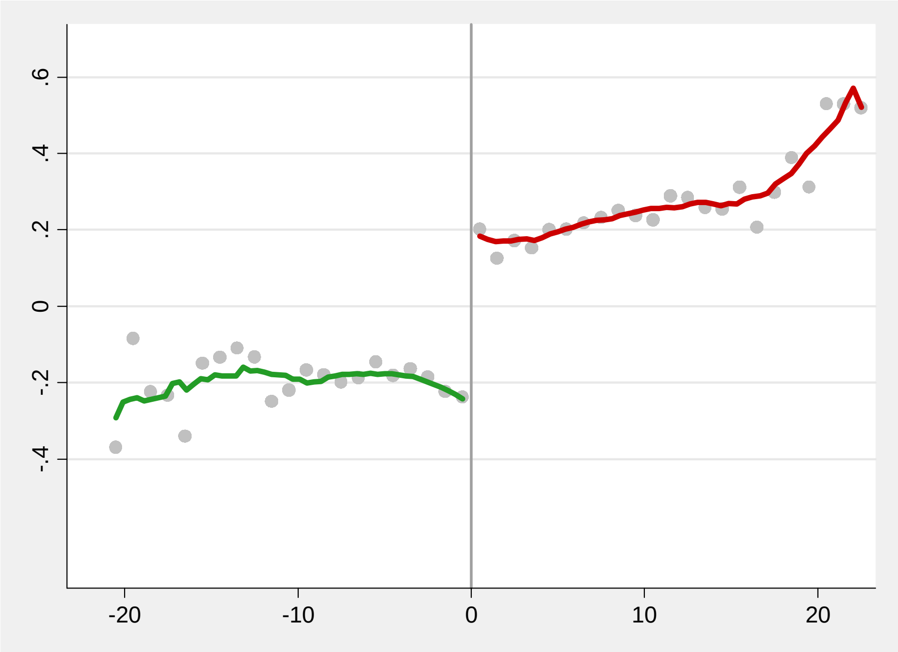
Notes: Entries are OLS estimates on a dummy of classrank, where classrank=0 for the younger class and classrank=1 for the older class. Marginal effects reported. For each variable a separate regression has been estimated. The data comes from the student questionnaire of PROEB. Heteroskedasticity robust standard errors are reported in parentheses. *** denotes significance at the 1% level.

TABLE 6
RD ESTIMATES OF MATH TEST SCORES

	Ranks from threshold in months				
	1 month	2 months	3 months	4 months	5 months
	Estimated discontinuity at threshold				
Quadratic	-56.574***	-54.578***	-59.044***	-57.193***	-59.182***
	(15.299)	(12.561)	(11.103)	(10.791)	(10.653)
Cubic	-55.477***	-54.467***	-59.560***	-57.188***	-58.416***
	(15.551)	(12.622)	(11.106)	(10.842)	(10.722)
Quadratic with full controls	-43.297***	-43.762***	-45.216***	-43.600***	-43.066***
	(15.673)	(12.446)	(11.259)	(10.980)	(10.675)
Cubic with full controls	-41.689**	-43.753***	-45.625***	-43.769***	-42.726***
	(16.299)	(12.45)	(11.274)	(11.031)	(10.749)
Number of observations	1688	3142	4547	5884	7223

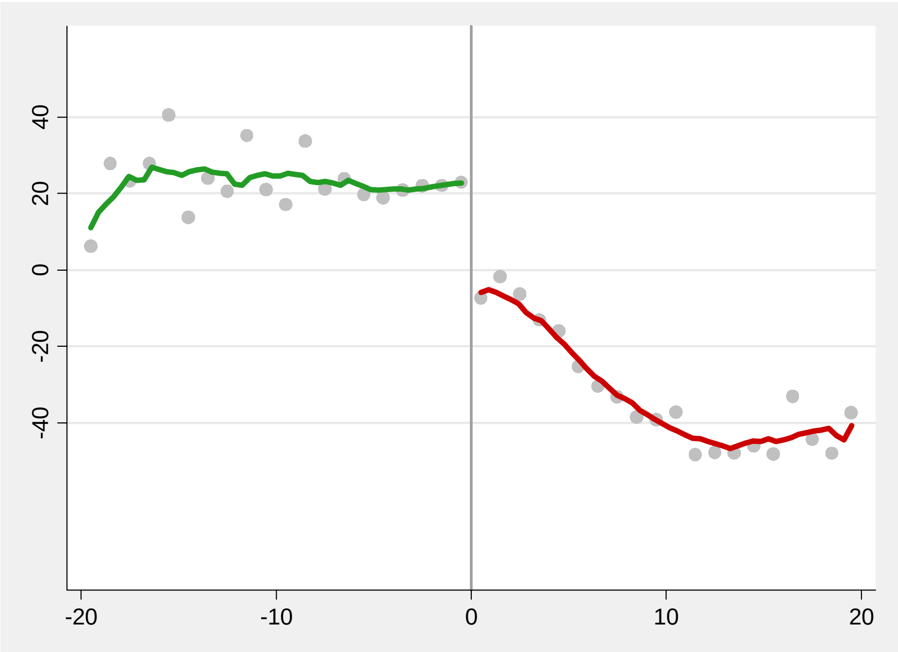
Notes: Dependent variable is math test score and entries are estimates of the discontinuity including different range of observations in terms of the age rank indicated by the column heading. Entries for row (1) are the estimated coefficients of the RD from models that include a quadratic polynomial in age for the different range of observations. Row (2) includes a cubic polynomial. Rows (3) and (4) additionally include the full set of controls as in column (4) of table 2. Heteroskedasticity robust standard errors are reported in parentheses. **, *** denote significance 5% and 1% level, respectively.

Figure 1: Class rank (treatment variable)



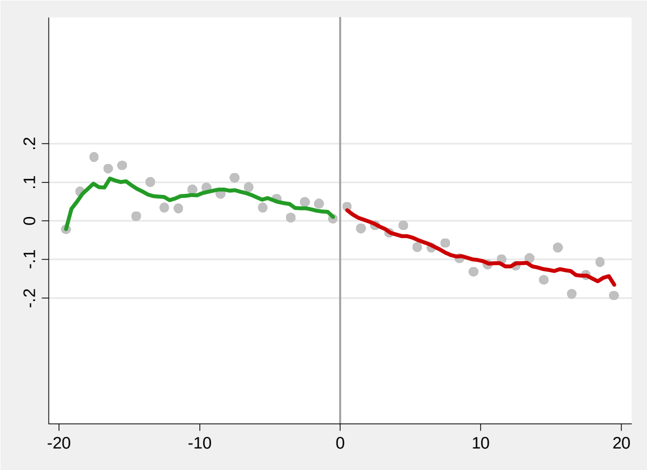
Notes: The graph plots local averages of the class rank of students as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 2: Math test score (outcome variable)



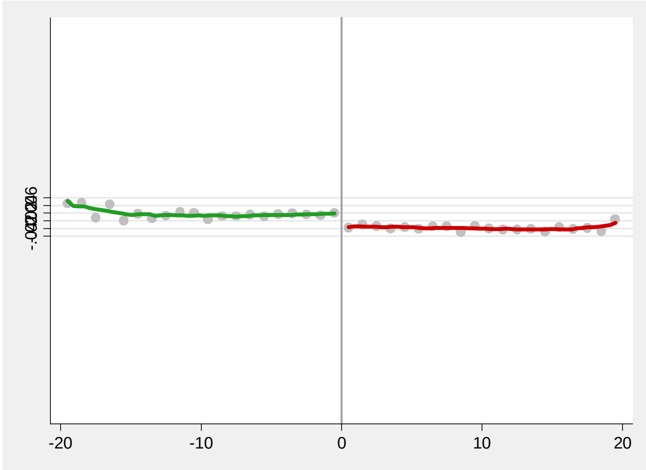
Notes: The graph plots local averages of math test scores as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 3a: Individual student sex composition



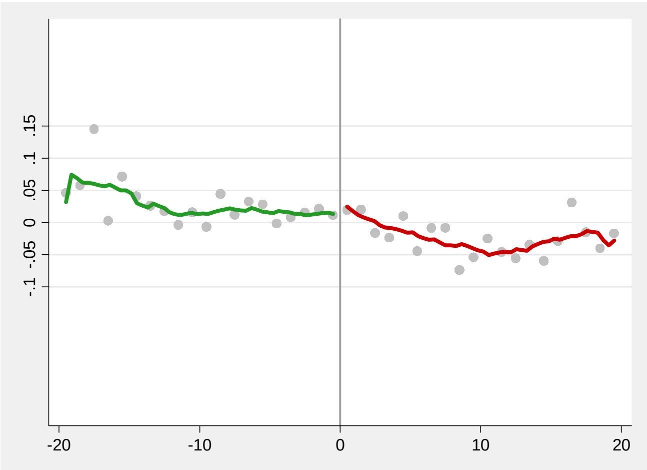
Notes: The graph plots local averages of the proportion of female students as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 3b: Peer student sex composition



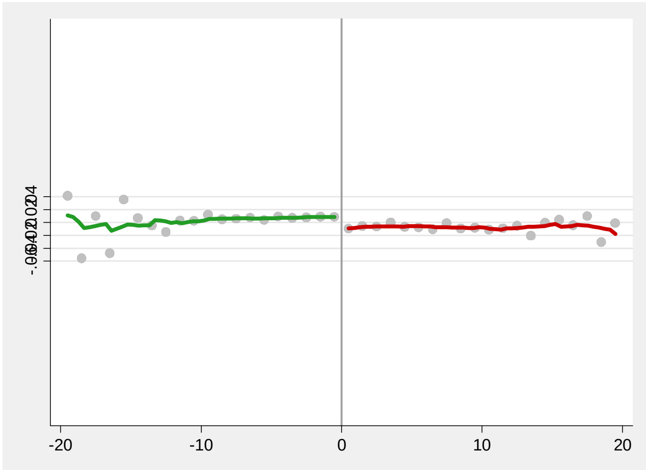
Notes: The graph plots local averages of the proportion of female students as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 4a: Individual proportion white students



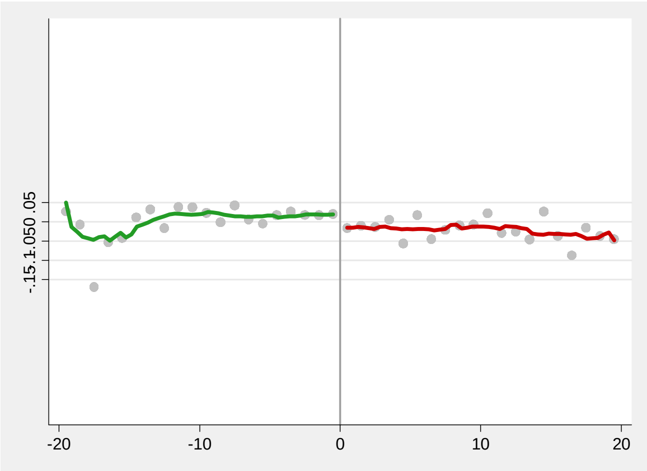
Notes: The graph plots local averages of the proportion of white students as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 4b: Peer proportion white students



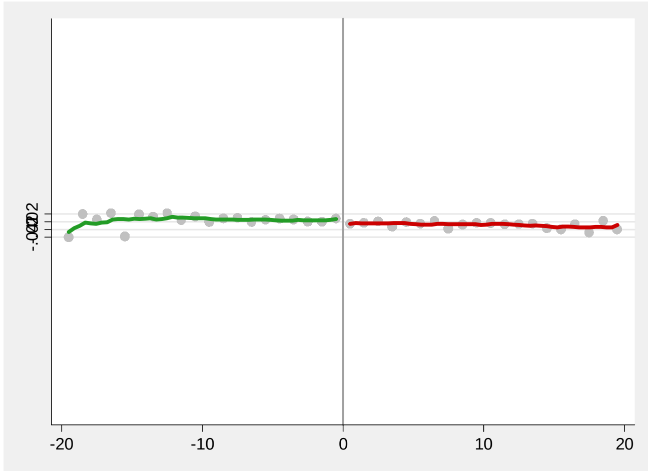
Notes: The graph plots local averages of the proportion of white students as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 5a: Individual proportion mixed students



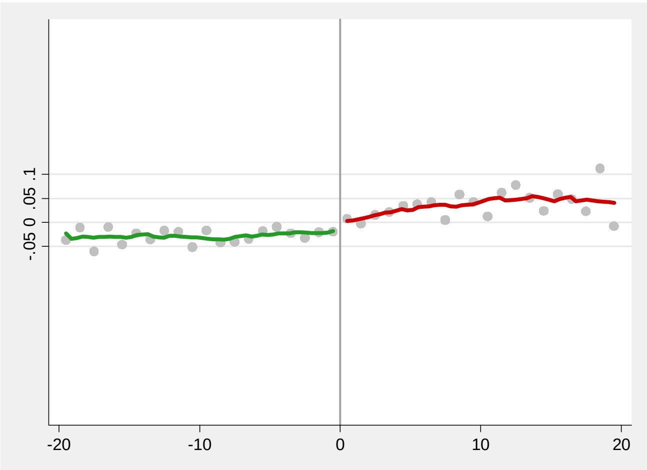
Notes: The graph plots local averages of the proportion of mixed students as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 5b: Peer proportion mixed students



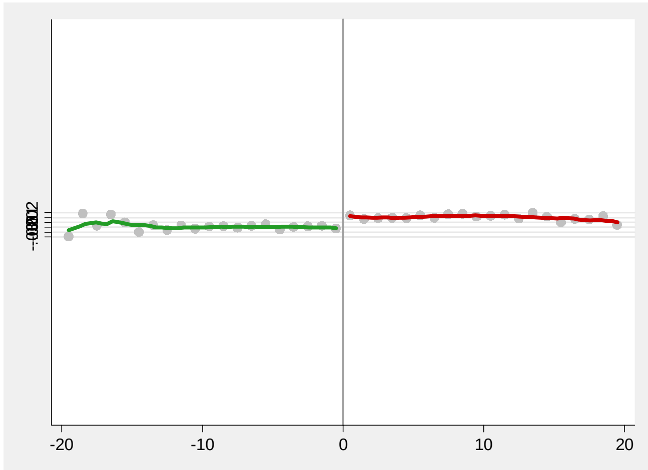
Notes: The graph plots local averages of the proportion of mixed students as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 6a: Individual proportion black students



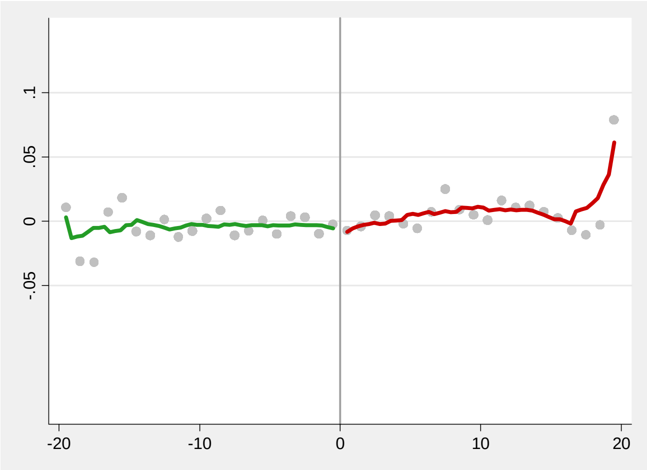
Notes: The graph plots local averages of the proportion of black students as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 6b: Peer proportion black students



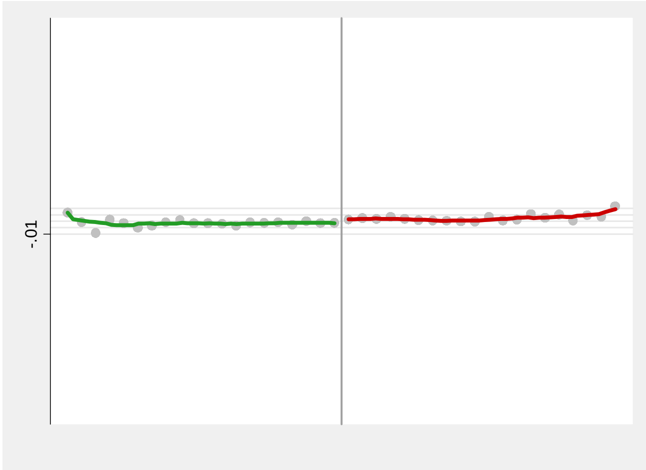
Notes: The graph plots local averages of the proportion of black students as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 7a: Individual proportion East-Asian students



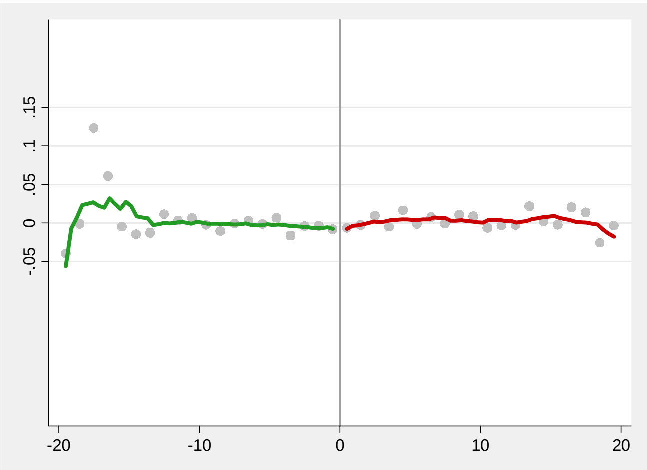
Notes: The graph plots local averages of the proportion of East-Asian students as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 7b: Peer proportion East-Asian students



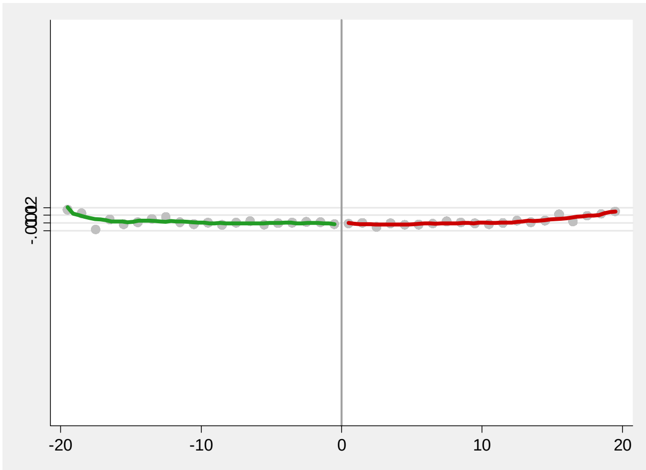
Notes: The graph plots local averages of the proportion of East-Asian students as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 8a: Individual proportion indigenous students



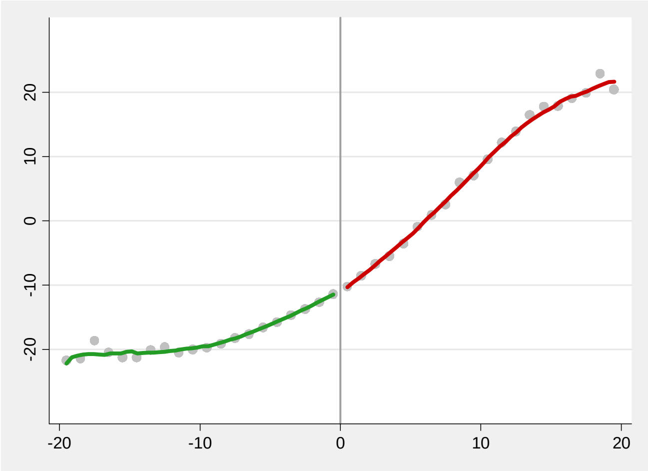
Notes: The graph plots local averages of the proportion of indigenous students as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 8b: Peer proportion indigenous students



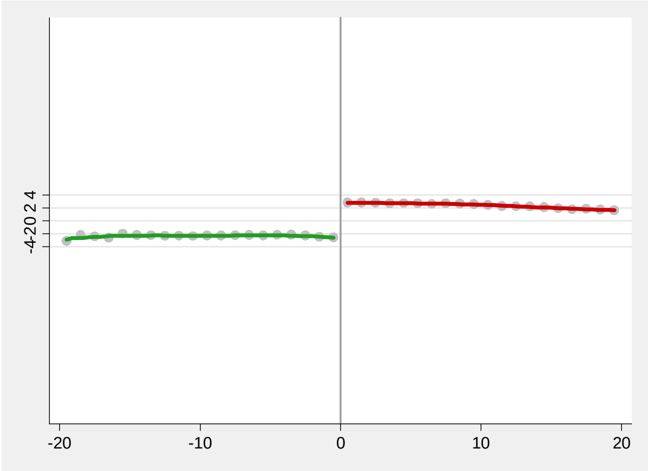
Notes: The graph plots local averages of the proportion of indigenous students as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 9a: Individual age



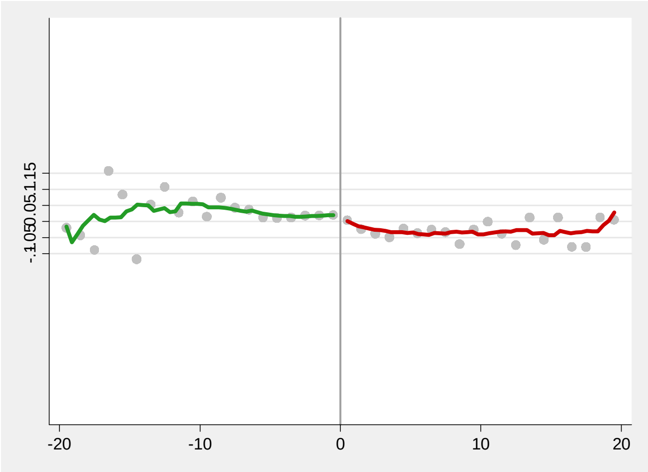
Notes: The graph plots local averages of student age (in months) as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 9b: Peer age



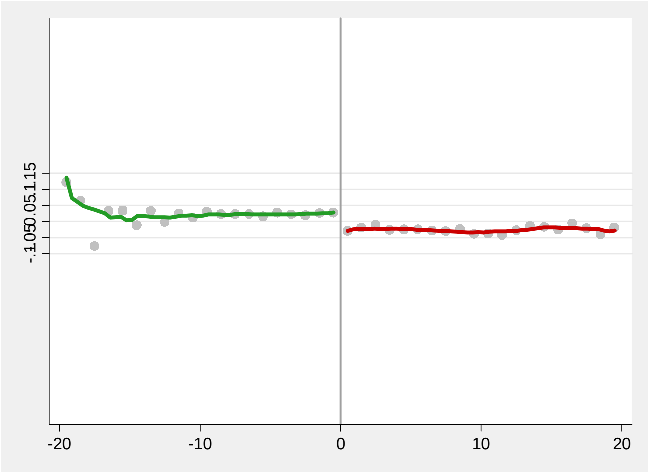
Notes: The graph plots local averages of peer student age (in months) as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 10a: Number of bathrooms in HH



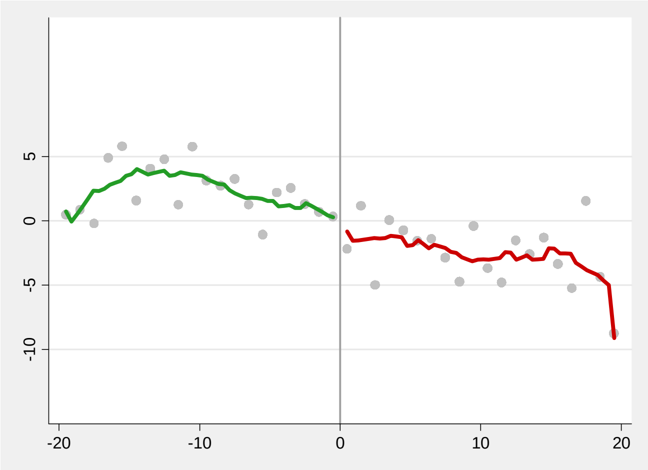
Notes: The graph plots local averages of the number of bathrooms in student HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 10b: Number of bathrooms in peer HH



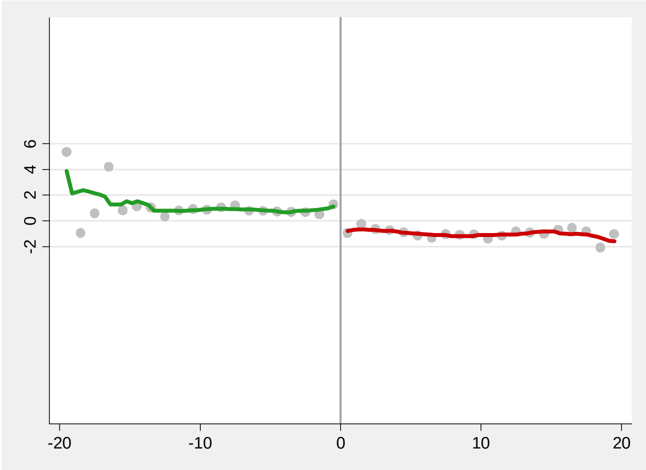
Notes: The graph plots local averages of the number of bathrooms in peer student HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 11a: Number of books in HH



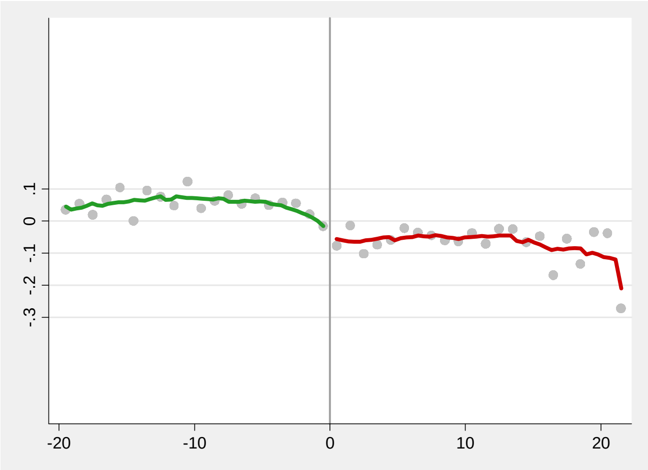
Notes: The graph plots local averages of the number of books in student HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 11b: Number of books in peer HH



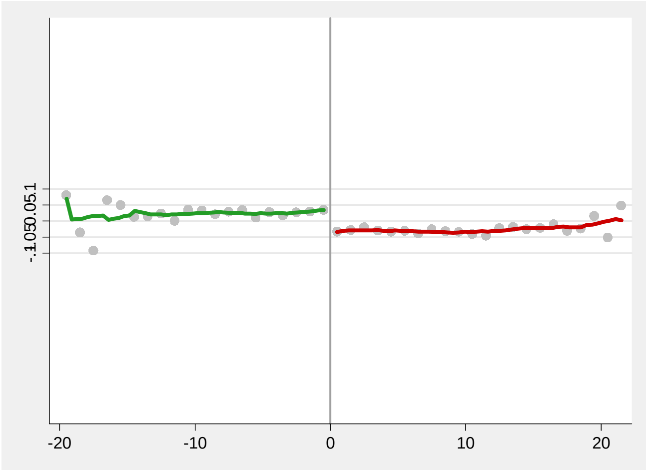
Notes: The graph plots local averages of the number of books in peer student HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 12a: Number of cars in HH



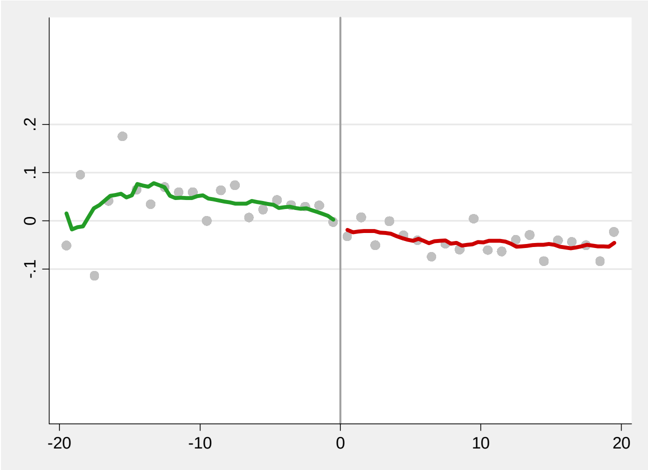
Notes: The graph plots local averages of the number of cars in student HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 12b: Number of cars in peer HH



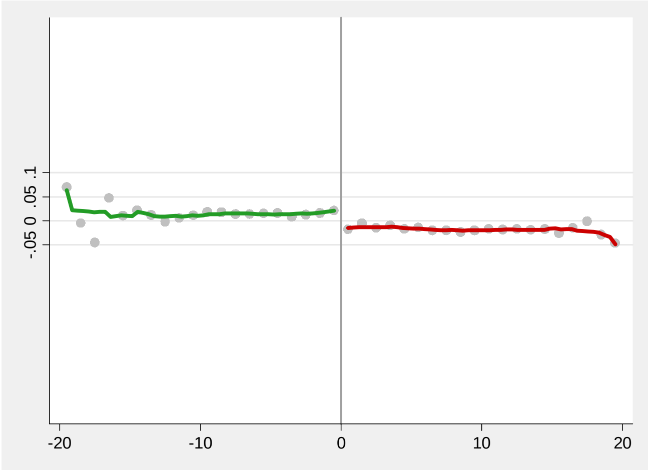
Notes: The graph plots local averages of the number of cars in peer HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 13a: Number of computers in HH



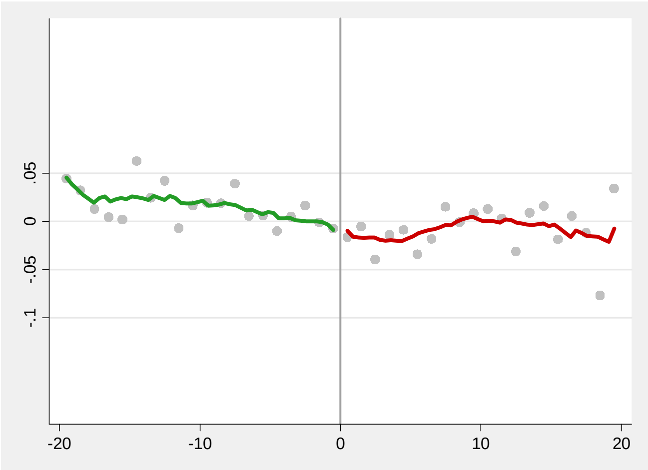
Notes: The graph plots local averages of the number of computers in student's HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 13b: Number of computers in peer HH



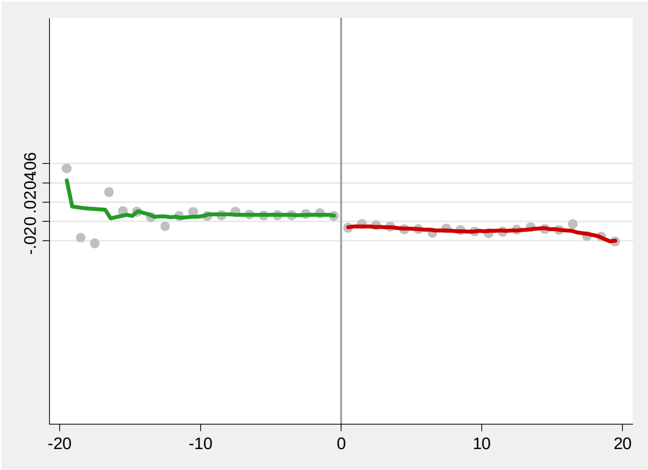
Notes: The graph plots local averages of the number of computers in peer HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 14a: Number of domestic workers in HH



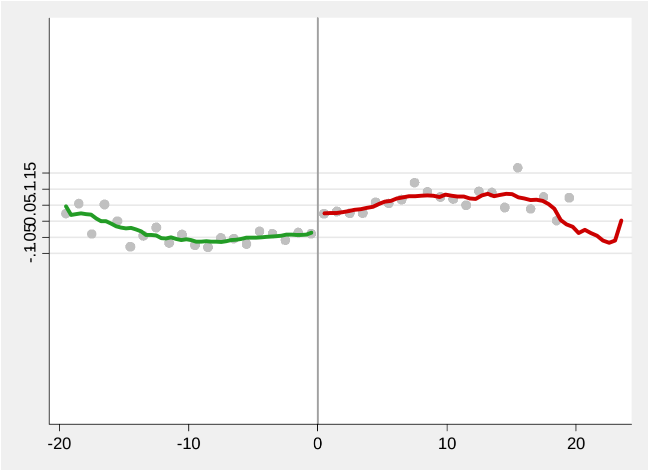
Notes: The graph plots local averages of the number of domestic workers employed by student HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 14b: Number of domestic workers in peer HH



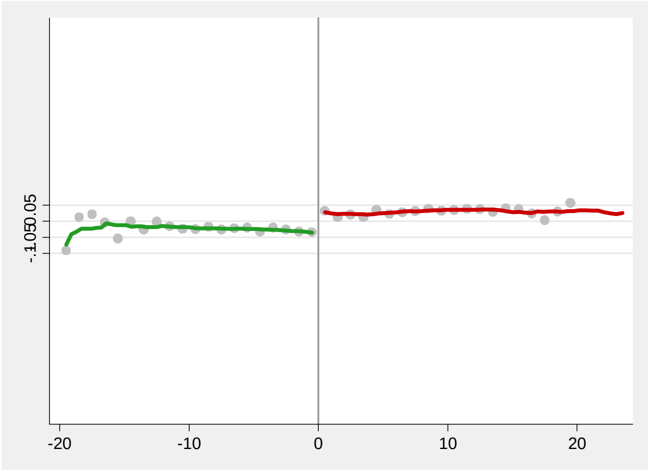
Notes: The graph plots local averages of the number of domestic workers employed in peer HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 15a: Proportion of HH with Bolsa Família



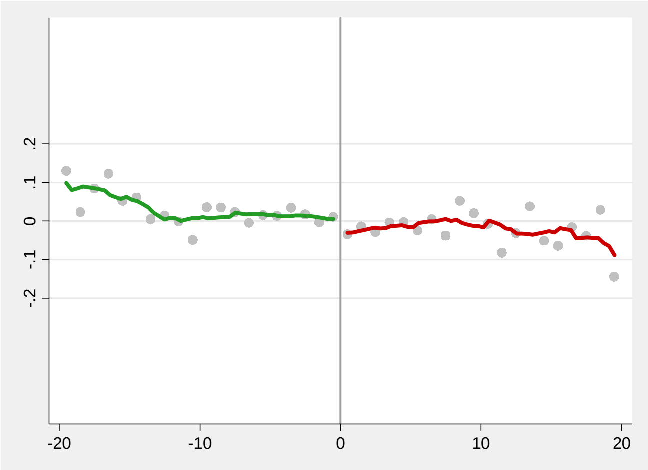
Notes: The graph plots local averages of the proportion of HH receiving Bolsa Família as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 15b: Prop. of peer HH with Bolsa Família



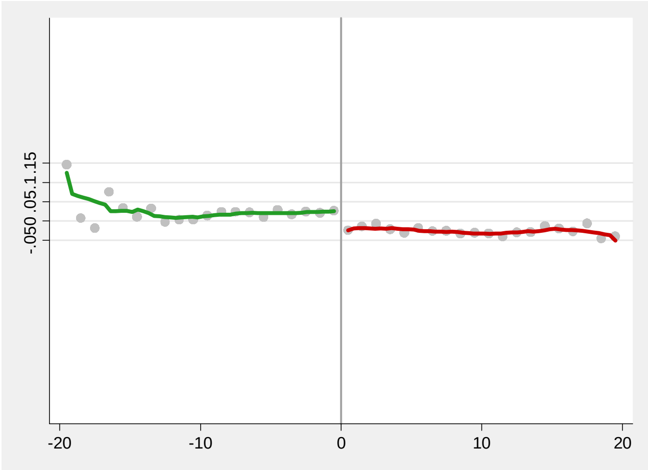
Notes: The graph plots local averages of the proportion of peer HH receiving Bolsa Família as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 16a: Number of DVD players in HH



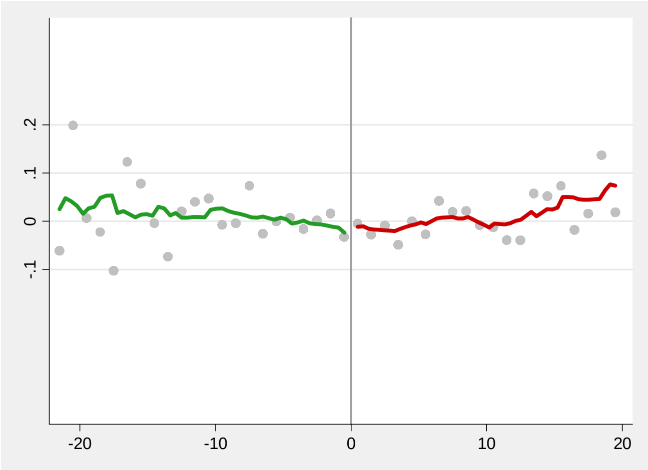
Notes: The graph plots local averages of the number of DVD players in student HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 16b: Number of DVD players in peer HH



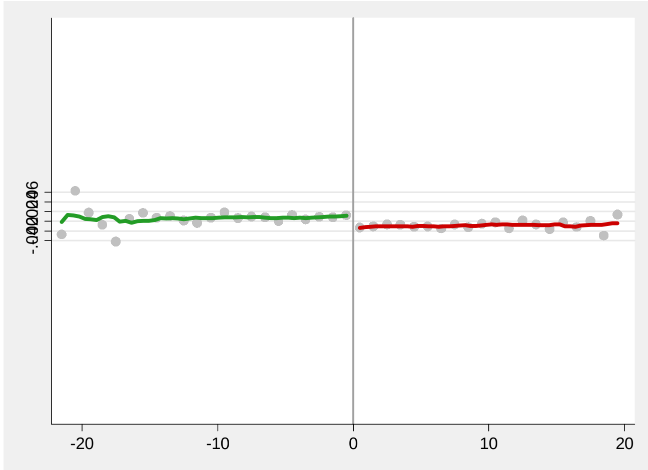
Notes: The graph plots local averages of the number of DVD players as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 17a: Number of freezers in HH



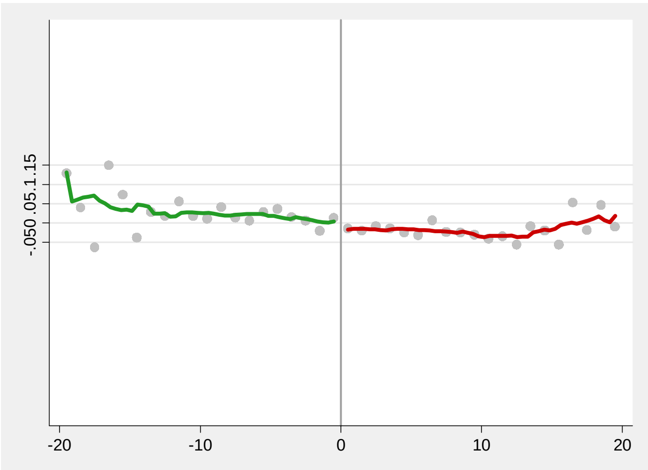
Notes: The graph plots local averages of the number of freezers in HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 17b: Number of freezers in peer HH



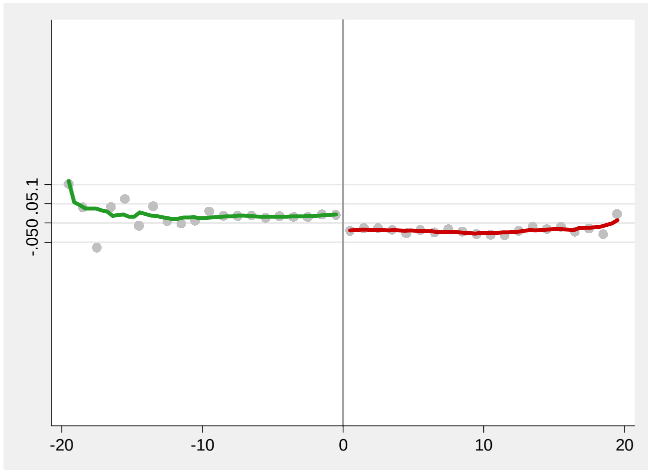
Notes: The graph plots local averages of the number of freezers in peer HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 18a: Number of fridges in HH



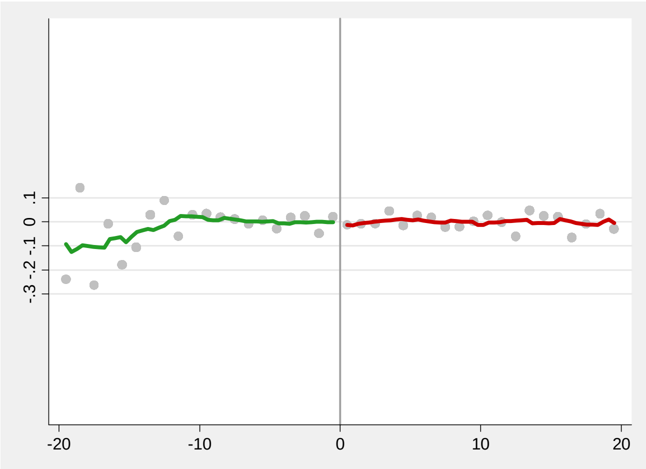
Notes: The graph plots local averages of the number of fridges in student HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 18b: Number of fridges in peer HH



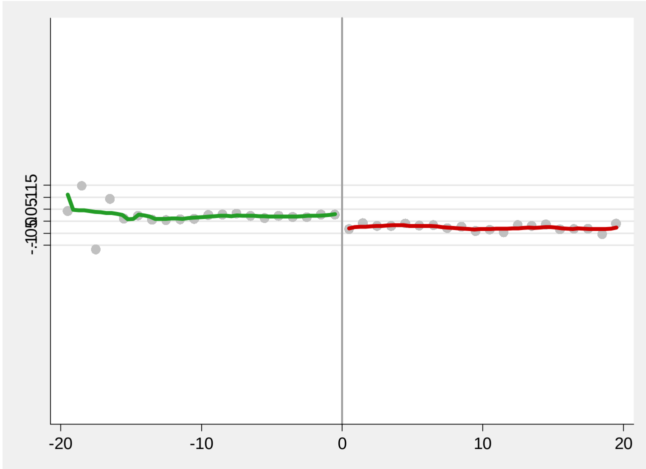
Notes: The graph plots local averages of the number of fridges in peer HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 19a: Number of radios in HH



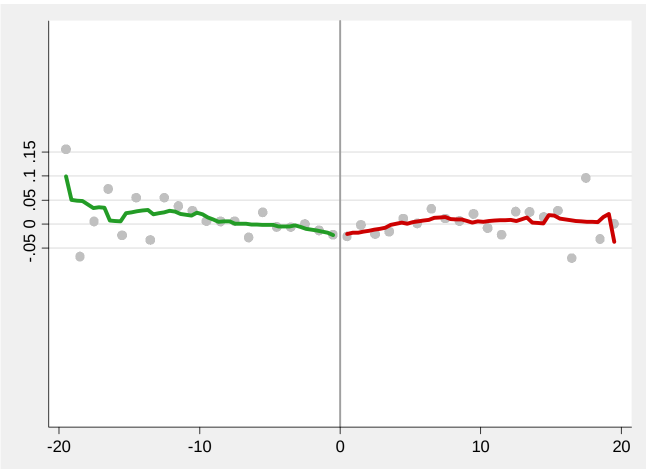
Notes: The graph plots local averages of the number of radios in student HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 19b: Number of radios in peer HH



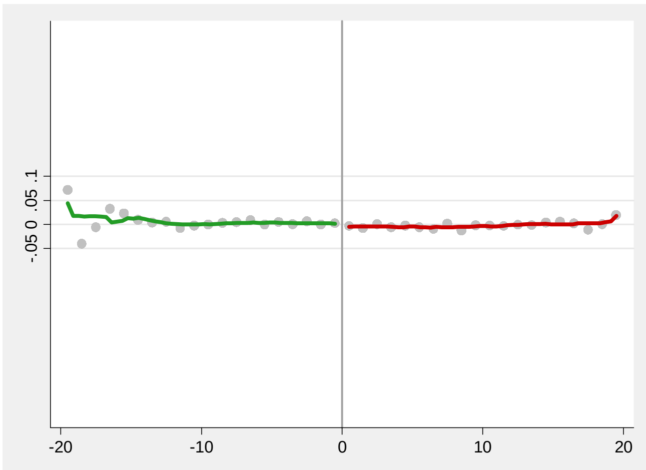
Notes: The graph plots local averages of the number of radios in peer HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 20a: Number of tumble dryers in HH



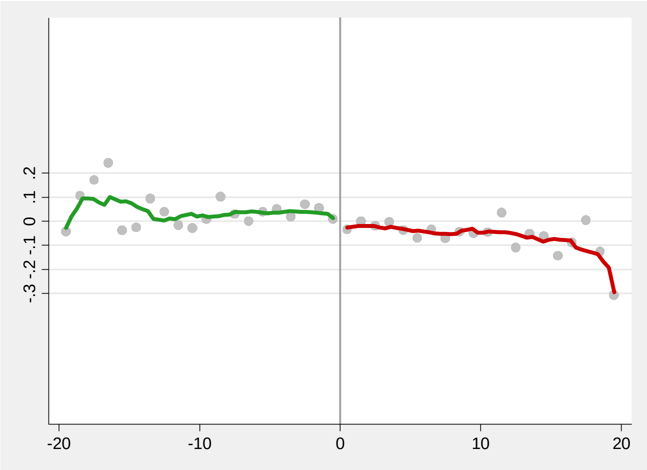
Notes: The graph plots local averages of the number of tumble dryers in student HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 20b: Number of tumble dryers in peer HH



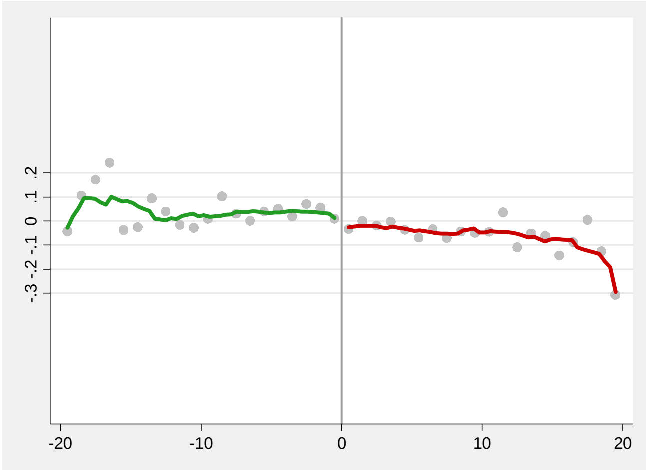
Notes: The graph plots local averages of the number of tumble dryers in peer HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 21a: Number of TV sets in HH



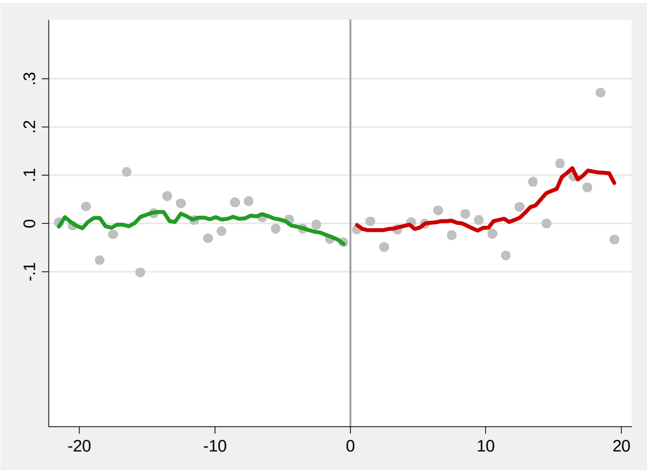
Notes: The graph plots local averages of the number of TV sets in student HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 21b: Number of TV sets in peer HH



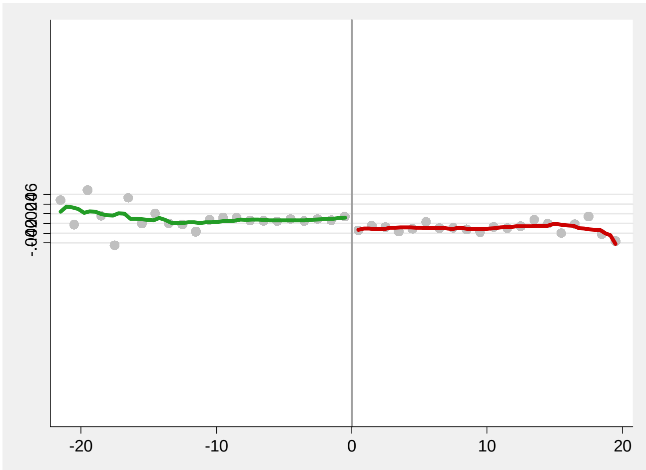
Notes: The graph plots local averages of the number of TV sets in peer HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 22a: Number of VCR's in HH



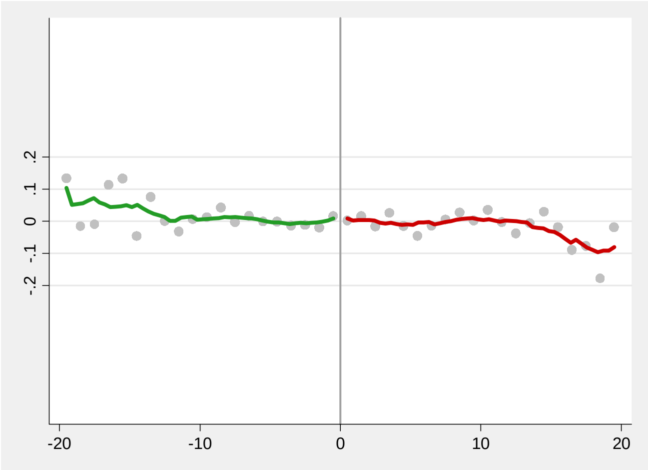
Notes: The graph plots local averages of the number of VCR's in student HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 22b: Number of VCR's in peer HH



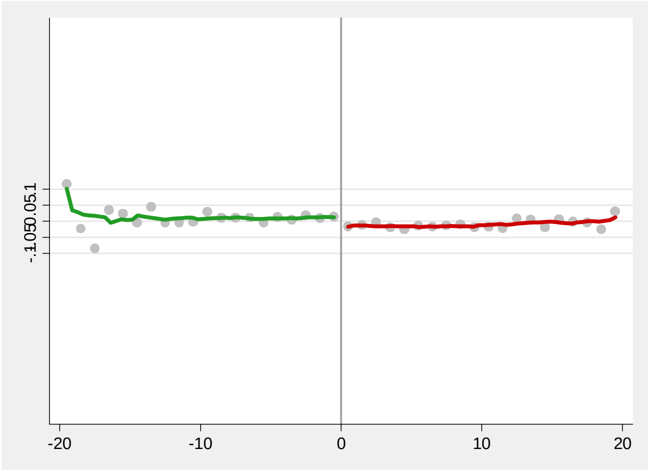
Notes: The graph plots local averages of the number of VCR's in peer HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 23a: Number of washing machines in HH



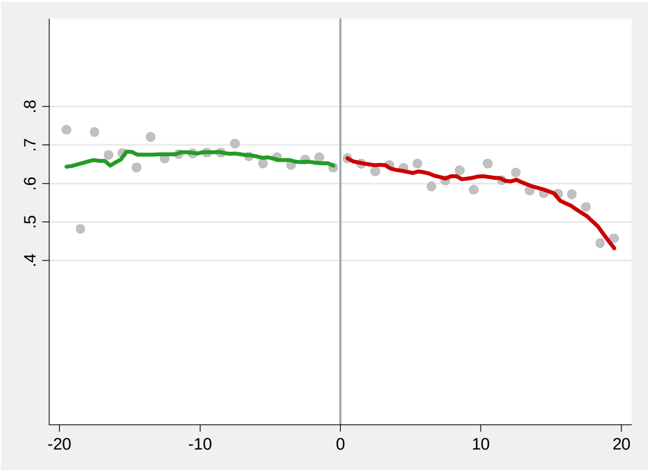
Notes: The graph plots local averages of the number of washing machines in student HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 23b: Number of washing machines in peer HH



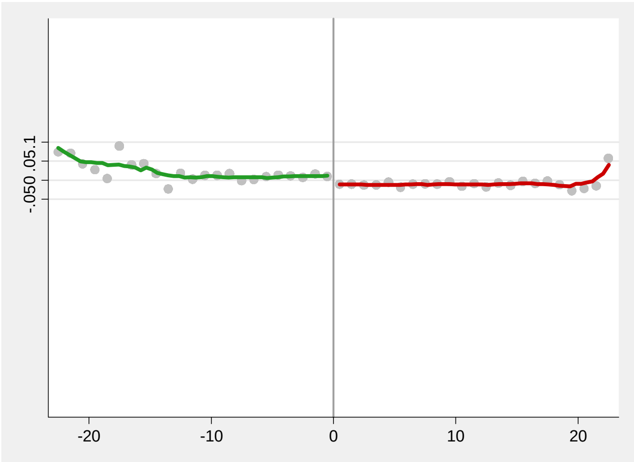
Notes: The graph plots local averages of the number of washing machines in peer HH as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 24a: Homework support from their parents



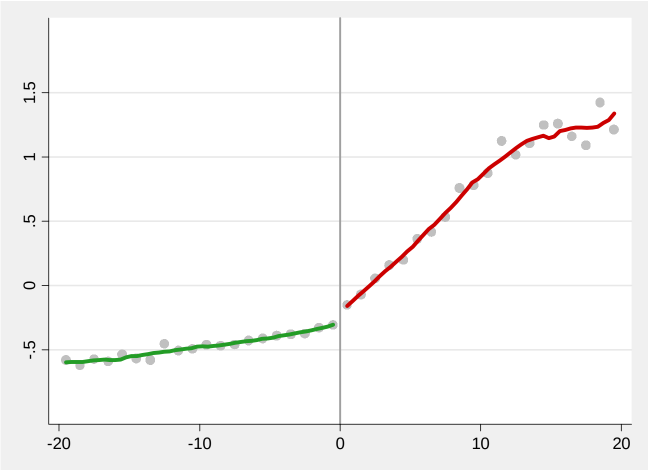
Notes: The graph plots local averages of the proportion of mixed students as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 24b: Homework support from their parents



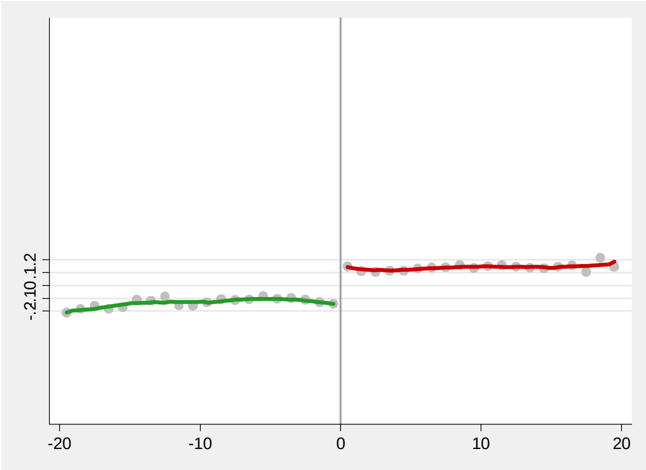
Notes: The graph plots local averages of the proportion of mixed students as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 25a: Grades repeated (in years)



Notes: The graph plots local averages of the proportion of mixed students as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

Figure 25b: Grades repeated of peers (in years)



Notes: The graph plots local averages of the proportion of mixed students as deviation from the mean school values according to the age ranking as distance of students from the cut-off point and local linear regression fits on both sides of the cut-off using a rectangular kernel with BW of 3 months.

A1 DATA ANNEX

The variable of individual age of students has been created based on three questions related to age of the test takers in the student questionnaire of PROEB. Students need to provide their age in years, their birth month and whether or not they already celebrated birthday in the current calendar year. This information together with the known test date of PROEB allows establishing the age of the children in years and months. Average age of students is 135.28 months, which is approximately 11.3 years. This is about ¾ of a year above the appropriate age at the end of 5^{th} grade. Average age in the younger classes is 131.15 months and 140.04 months in the older class. The standard deviation of age in the cohort at 5^{th} grade is 12.09 months. The distribution of age in the two classes differs quite considerably with a standard deviation of age in the younger classes of 10.02 month and 14.16 in the older classes. The histograms of figures A1 and figure A2 show the different distribution of age in the two classes. Both distributions are positively skewed, with the mass of the distribution concentrated to the left. This is due to age being naturally limited at the bottom with a minimum enrolment age of 5½, and the upper limit in age (maximum observed age is 15 years, which is almost 4 years above average age and 5 ½ years above the possible youngest age).

The substantial age-grade distortion in the student cohort can mostly be attributed to grade repetition of students. Every year repeated by a student adds to the given age variation based on the distribution of birth dates and the enrolment cut-off point at first grade. With 20% of students having repeated one year, 9% having repeated twice and 4% having repeated three or more times, repetition accounts almost completely for the age-grade distortion observed in the data (grade repetition accounts for approximately half a year in mean student age). The remainder is likely due to some late enrolment and school dropout with re-enrolment or school changes of students with reassignment at a lower grade. Unfortunately, I do not have available information on enrolment age for the cohort of interest. From the school census 2007 that contains information on age for individual students for first grade, I can infer that late enrolment was responsible for about 1.8 months, which is likely similar to late enrolment in the cohort of consideration that has enrolled 4 years earlier.

A2 ANNEX TO ORGANIZATION OF STUDENTS INTO CLASSROOMS

This annex gives some more details about the allocation mechanism of students into classes.

The initial allocation of students is completed at the beginning of first grade prior to school staff observing ability, and with age of the students being readily observable, allocating students according to age is an accessible way of allocating students into classes. As class composition of students is stable for at least the first five years of schooling, only migration between schools, drop-out and grade repetition affects the composition of the classes. Classes in which students have been sorted to make them heterogeneous in age, have an average age of 133.8 for the older classes and 130.0 months for the younger classes and a t-test for the equality of the mean between the two classes reveals that there is no statistically significant difference in mean age. Schools, in which students are allocated to classes according to their relative age, have a mean age of 131.2 and 140.0 months, respectively.[33]

To sort students to form homogeneous age classes, the school administration ranks students according their age. With more than one class and a maximum class size rule of N students, the first class is formed by assigning students starting with the youngest student, being followed by the next youngest and so on, until the class size cap of N has been reached and the student N+1 is assigned to the second class etc. This rule is similar to the class-size function outlined by Angrist and Lavy in their paper on Maimonides rule (1999).[34] Treating cohort size e_s as exogenous and with a maximum class-size of 25 students in Minas Gerais[35] and under the assumption that the cohorts are divided into classes of equal size, class-size f_{sc} in school s and class c is a function of initial cohort size e_s, and f_{sc} is given as

$$f_{sc} = e_s/(\text{int}\ ((e_s - 1)/25)+1)$$

where the function int(e_s -1) gives the largest integer smaller or equal to e_s-1 (Angrist and Lavy 1999). If cohort size is below the exact multiples of N, the maximum class-size rule does not bind and the cohort can be divided in an arbitrary way, as long as the maximum class size is not met. Although there is a maximum class size rule in Minas Gerais, there is widespread circumvention of this rule. This is obvious considering the very high cost from creating an additional class for a cohort just above multiples of the class size cap. With a given example entry cohort of 54 students strictly following the class size cap, average class size would turn out to be 18 in three classes compared to 27. It may often not be possible in particular for small schools with two classes per grade to accommodate an additional class in the given school space and employ additional teaching staff, so that schools with a cohort size slightly exceeding the class size caps, are permitted to do so.

Although identifying assumptions for the regression-discontinuity are not violated by a non-random choice of the allocation rule of students into classes as identification relies on a local discontinuity within schools, it may be helpful to understand what drives the headmasters and school administrators to choose grouping students into classes at first grade in a particular manner. For that purpose, I have estimated a linear probability model, where the dependent variable is a binary variable with a value=0 if students are grouped heterogeneously in age and a value=1 if students are grouped homogeneously in age into classes. I use rich information on physical school characteristics and headmaster, teacher and mean student

[33] See table A3.

[34] See Urquiola and Verhoogen (2009) with a discussion on the validity of the approach by Angrist and Lavy (1999) due to strategic behaviour of schools close to the multiples of the class size cap in the case of private schools.

[35] Law No 16.056 from 24th April 2006 establishes a maximum class-size for the first 5 years of fundamental schooling of 25 students; exceptions to that rule are only permitted in cases exceptional circumstances or transitory situations.

characteristics as exogenous variables to infer about whether these are relevant determinants for the decision on the allocation rule. Specifically I estimate the following linear model: $Y = \beta_0 + \beta_1 S + \beta_2 D + \beta_3 T + \beta_4 P + u$, where Y is a binary outcome variable of choosing to sort students according to age making classrooms heterogeneous in age classrooms (Y=0) or homogenous in age (Y=1), S denotes school characteristics, D headmaster characteristics, T teacher characteristics, P mean characteristics of pupils in the cohort and u an idiosyncratic error term. Table A2 reports the estimated coefficients from the model; only few variables show a statistically significant effect on the grouping choice: absolute cohort size, the number of books in the parental household, the dummy on the existence of a secretariat and the participation in a computer literacy programme, the number of Pentium computers in the school and the professional experience of headmasters in years. With a larger cohort size, administrators tend to chose heterogeneous age sorting and with a student body with better socio-economic background (proxied by mean number of books in the HH) headmasters tend towards homogeneous age sorting. Also, headmasters with more years of experience tend to prefer homogeneous age sorting, but the size of the effect is negligible. Other coefficients are only marginally statistically significant at the 10% level. The selection of specific sorting schemes by the school administration nevertheless does not affect the identifying assumptions of the empirical strategy.

TABLE A1
MEANS AND PROPORTIONS SCHOOL AND HEADMASTER CHARACTERISTICS

School physical characteristics			
Means	Permanent class rooms	10.25	(0.19)
	Number of total staff	46.11	(1.15)
	Number of teaching staff	26.98	(0.98)
	Computers for students	10.11	(0.40)
	School books for 5th grade	290.13	(20.59)
	Class size	26.66	(0.01)
Proportions	Urban school	0.91	(0.02)
	State school	0.55	(0.03)
	Municipal school	0.45	(0.03)
	Filtered water	0.99	(0.01)
	Building shared with other school	0.10	(0.02)
	Directorial office	0.90	(0.02)
	Faculty room	0.84	(0.02)
	School library	0.83	(0.02)
	Video collection	0.36	(0.01)
	TV room	0.98	(0.01)
	Video player	0.90	(0.02)
	DVD player	0.85	(0.02)
	Copy machine	0.37	(0.02)
	Kitchen	0.93	(0.01)
	Internet connectivity	0.59	(0.03)
	School canteen	0.54	(0.04)
	Computer laboratory	0.35	(0.02)
	Science laboratory	0.11	(0.02)
	Facilities for disabled children	0.82	(0.02)
	Public water supply	0.95	(0.01)
	Public energy supply	1.00	(0.00)
	Public sewage	0.83	(0.02)
	Waste collection	0.91	(0.01)
	Minimum income programme	0.98	(0.01)
	TV escola (school TV programme)	0.44	(0.04)
	Public school transport	0.80	(0.03)
	School lunch	0.95	(0.01)
Director characteristics			
Sex	Female	0.86	(0.02)
Race	White	0.43	(0.03)
	Mixed	0.42	(0.03)
	Other	0.07	(0.01)
	Mean age	43.1	(0.05)
highest edu. level	Secondary education	0.05	(0.12)
	Higher education – ped. degree	0.32	(0.03)
	Higher education – math	0.43	(0.03)

Higher education – literature	0.05	(0.01)
Higher education – other	0.15	(0.02)
Salary (in R$)	1635.49	(38.85)
Years of experience in education	18.09	(0.21)
Years of experience at this school	6.21	(0.24)
Years of experience as director	6.95	(0.26)
Participation in continued training	0.11	(0.02)

Notes: Data for the physical school characteristics comes from the annual Brazilian school census, headmaster characteristics come from the 2007 wave of PROEB.

TABLE A 2

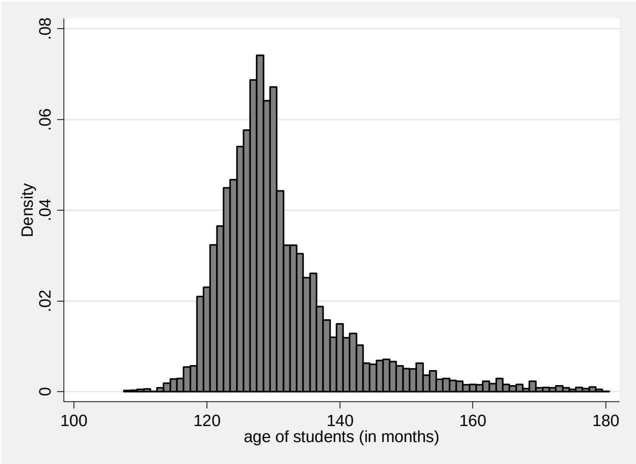
LINEAR PROBABILITY MODEL

		coefficient	s.e.
SCHOOL PHYSICAL CHARACTERISTICS			
	Urban school	0.00	(0.10)
	State school	0.05	(0.08)
	Number of permanent class rooms	0.00	(0.02)
	Number of provisory class rooms	-0.09	(0.07)
	Number of class rooms	0.02	(0.02)
	Total number of staff	0.01	(0.01)
	Total number of teachers	-0.01	(0.01)
	Size of cohort	-0.01***	(0.00)
	Principal office	-0.10	(0.10)
	Secretarial office	-0.28**	(0.13)
	Faculty room	0.11	(0.10)
	Video collection	-0.22	(0.20)
	TV room	0.00	(0.08)
	Kitchen	-0.03	(0.09)
	School canteen	-0.11	(0.07)
	Computer lab	-0.05	(0.09)
	Science lab	-0.07	(0.13)
	Public energy supply	-0.19	(0.24)
	Public water supply	0.09	(0.18)
	Public sewage	0.04	(0.08)
	Minimum Income Programme	0.21	(0.15)
	TV escola	-0.08	(0.07)
	Project Saude	0.10	(0.16)
	Computer Literacy Programme	0.21**	(0.09)
	Other federal programmes	-0.09	(0.09)
	Other state programmes	0.08	(0.10)
	Other municipal programmes	-0.06	(0.06)
	Public school transport	-0.07	(0.08)
	Number of video player	-0.02	(0.04)
	Number of TV sets	0.00	(0.04)
	Number of overhead projectors	-0.01	(0.06)
	Number of printers	-0.02	(0.03)
	Number of sound systems	0.02	(0.02)
	Number of Pentium computers	0.02**	(0.01)
	Number of 386/486 computers	0.02	(0.01)
HEADMASTER CHARACTERISTICS			
	Headmaster male	-0.12*	(0.06)
	Headmaster age	0.0002*	(0.00)
Headmaster highest education	High school	-0.09	(0.11)

	Higher edu - pedagogic degree	-0.05	(0.11)
	Higher edu - normal	-0.13	(0.11)
	Higher edu & teaching qualification	-0.09	(0.12)
	Higher edu – other	-0.06	(0.14)
	Experience in years as headmaster	0.0003**	(0.00)
	Experience in years in education	-0.0006*	(0.00)
	Continued training	0.02	(0.06)
<u>TEACHER CHARACTERISTICS</u>			
Teacher characteristics	Proportion male	0.05	(0.12)
	Higher education	0.07	(0.12)
Teacher highest edu.	High school	-0.13	(0.11)
	Higher edu. - pedagogic degree	0.13*	(0.07)
	Higher edu. - normal	-0.16	(0.12)
	Higher edu. & teaching qualification	-0.01	(0.06)
	Higher edu. – other	0.16	(0.13)
	Participation in teacher training	-0.01	(0.06)
<u>STUDENTS CHARACTERISTICS</u>			
Student characteristics	Proportion Bolsa Família	0.04	(0.20)
	Mean books	0.01***	(0.00)
	Mean homework help	-0.01	(0.16)
	Proportion female	0.16	(0.25)
	Mean HH with domestic worker	0.43	(0.33)
	Proportion white	-0.13	(0.16)
	Mean automobiles	-0.11	(0.13)
	Mean computers	-0.14	(0.26)
	Mean times teacher absent	0.02	(0.10)
	Constant	1.16***	(0.38)
	Observations	363	
	R-squared	0.29	

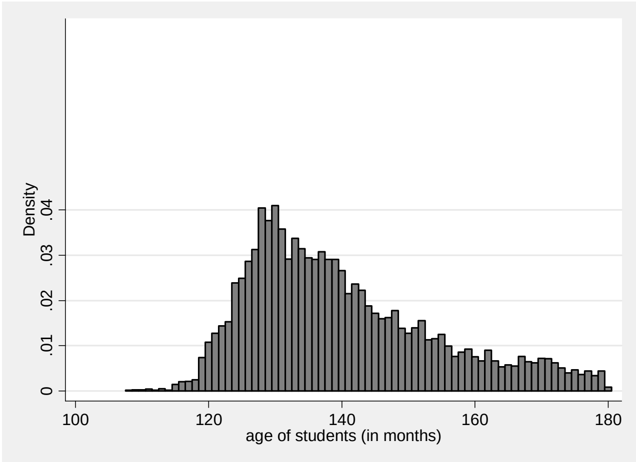
Notes: Standard errors in parenthesis, * significant at 10% level; ** significant at 5% level; *** significant at 1%

Figure A1: Age histogram for class 1.



Notes: The graph plots the density for age of students for class 1 (younger class), age is reported in months.

Figure A2: Age histogram for class 2



Notes: The graph plots the density for age of students for class 2 (older class), age is reported in months.